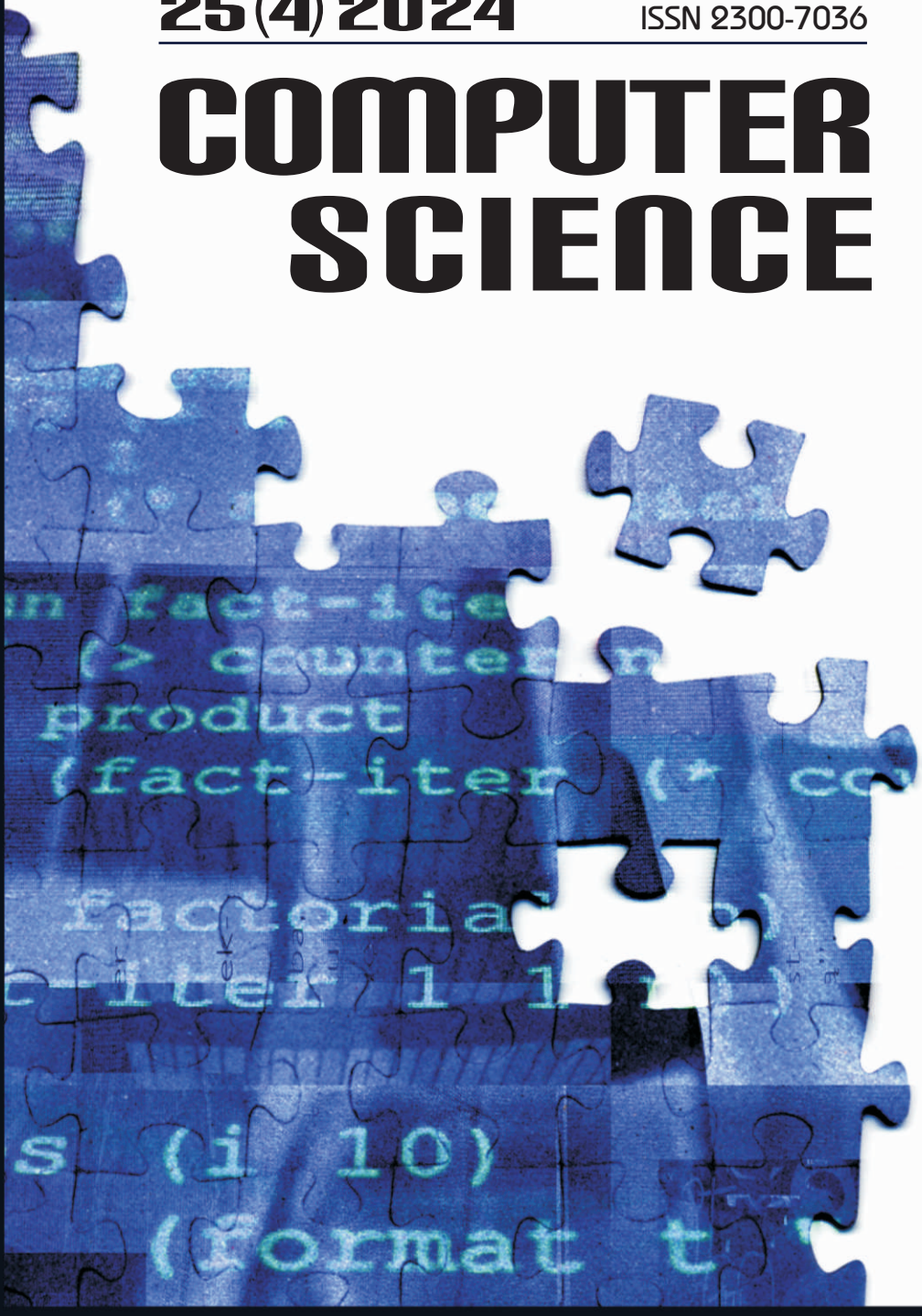
25(4) 2024 ISSN 2300-7036

COMPUTER SCIENCE



QUARTERLY OF AGH UNIVERSITY OF SCIENCE AND TECHNOLOGY

AGH UNIVERSITY PRESS KRAKOW 2024

25(4) 2024 ISSN 2300-7036

COMPUTER SCIENCE

AGH UNIVERSITY PRESS KRAKOW 2024

COMPUTER SCIENCE is published by AGH University Press, Krakow, Poland.

The papers presented in COMPUTER SCIENCE have been accepted by the reviewers selected by the editors of the journal.

Technical Editor
Magdalena Grzech

Ghostwriting prevention
Łukasz Faber

Statistical Correction
Anna Barańska

Linguistic Correction
Bret Spainhour

Cover Design
Anna Sadowska

Typesetting and Desktop Publishing
Marek Karkula

ISSN 2300-7036
DOI: https://doi.org/10.7494/csci

Wydawnictwa AGH (AGH University Press)
al. A. Mickiewicza 30, 30-059 Kraków, Poland
tel. +48 12 617 32 28, +48 12 636 40 38
e-mail: redakcja@wydawnictwoagh.pl
https://www.wydawnictwa.agh.edu.pl

Computer Science • 25(4) 2024 https://doi.org/10.7494/csci.2024.25.4.6289

Jacek Długopolski
Marcin Sadowski
Wawrzyniec Suleja

A PHYSICAL MODEL OF QUANTUM BIT BEHAVIOR BASED ON A PROGRAMMABLE FPGA INTEGRATED CIRCUIT

Abstract *The rapidly developing field of quantum computing and the ongoing lack of widely available quantum computers create the need for scientists to build their simulators. However, mathematical simulation of such circuits usually ignores many aspects and problems found in real quantum systems. In this article, the authors describe a quantum bit emulator based on FPGA integrated circuits. In this case, FPGA technology provides real-time massive parallelism of the modeled physical phenomena. The modeled QUBIT is represented using a Bloch sphere. Its quantum state is set and modified only by precise pulses of an electrical signal, and with the help of similar pulses, it manifests its current state in real time. The constructed QUBIT was additionally equipped with decoherence mechanisms and with circuits that intentionally respond to internal and external noises that distort its current quantum state. This article presents and discusses how such a physically built emulator works.*

Keywords quantum, qubit, FPGA, parallel, real time

Citation Computer Science 25(4) 2024: 499–519

1. Introduction

Currently, quantum physics is increasingly entering the world of information processing. The specific features of quantum phenomena (such as superposition, tunneling, or entanglement) mean that data processing carried out on the basis of quantum bits (qubits) allow achieving useful computational results in a time that would not be possible to achieve using classical methods.

However, quantum systems are still very difficult to build and maintain, so they still may not be widely available in the near future. Therefore, the idea of building a practical model of the behavior of the basic unit of a quantum computer (qubit) seems justified. After equipping such a model with a physical control system using electrical impulses, as well as a natural susceptibility to internal and external noise and sensitivity to decoherence, we will obtain a useful tool for researching and better emulating real quantum systems. By combining several such emulators, you can create more complex quantum systems and test sample quantum computing and noise reduction algorithms on them.

Effective, fault-tolerant quantum computing requires two-or-more-qubit gates with the highest possible fidelity. The recommended infidelity in the literature is below 10^{-4} [17]. Achieving such a level of fidelity in two-qubit gates implemented in an ion trap, such as the SWAP gate and the Mølmer-Sørensen gate, even in noise-free conditions, requires analysis not only of the internal states of the ions but also of the most relevant part of the phonon space. The authors hypothesize that complex emulations of this kind can be more efficiently conducted using parallel and real-time models based on FPGA integrated circuits, which may in the future enable more effective identification of both coherent and stochastic error sources. In addition, quantum processors also rely on classical electronic controllers to manipulate and read out the state of qubits. As the performance of the quantum processor improves, imperfections in the classical controller can become the performance bottleneck for the whole quantum system. To prevent such limitations, a systematic study of the impact of classical electrical signals on qubit fidelity is needed. To have an easy way to do such a study, we need a model close enough to real quantum systems. By controlling the qubit model only with real physical signals (pulses), it is possible to more accurately reproduce the control conditions of real quantum systems. Our proposal described in this article tries to meet all these needs.

This article describes the practical implementation of just such a model, based on a programmable FPGA integrated circuit. In addition to the qubit model (QUBIT), the article also details the control and reading module (GENMET), designed to generate appropriate electrical control pulses, as well as to read electrical pulses generated by the QUBIT itself, through which it informs the environment about its current state.

The article will also describe the idea of communication with such a single QUBIT and the method used to implement its susceptibility to decoherence and to internal and externally generated noise.

Next, the results and practical tests of the built emulator using the dedicated GENMET module will be described. The user can, of course, use other generators and measurement systems instead of GENMET. Finally, a brief description of possible applications will be presented.

2. Related works

At the time of writing this article, the authors had not found another similar solution in the world. Of course, many quantum simulators and emulators using mathematical apparatus are being created [1–4, 6–8, 10–15, 18], but they do not give the user the opportunity to encounter the physical problems experienced when working on building real quantum computers. The solution proposed in this article necessitates precise formation and measurement of electrical impulses that are susceptible to unpredictable physical disturbances. This provides a unique opportunity to encounter these types of problems. There are, of course, various solutions to mathematically generate artificial noise in quantum circuit simulators. An example of such a system is, for example: "Qiskit Aer – the high-performance quantum computing simulators with realistic noise models" [9]. This system, however, uses a mathematical formalism to add noise. Our system uses physical noise, which can be natural and come from within the FPGA chip itself, or it can come from an external physical function generator whose generated waveforms are also not completely immune to noise. Since we feed both control pulses signals and noise signals to our QUBIT via ordinary electrical cables, the signals flowing in them can be further naturally disturbed by external electromagnetic fields, as happens in real quantum systems under construction today. In this respect, it gives our system described here an undoubted advantage over other simulators available. The user here is getting very close to what she or he will be dealing with when operating non-ideal real quantum systems.

3. The idea of a quantum bit

A quantum bit (qubit) is different from a classical bit. While a classical bit is the basic smallest unit of information that can represent one of two possible classical states: '0' or '1', according to formula:

$$bit = \begin{cases} '0' \\ '1' \end{cases}$$

a qubit can take on these two classical values simultaneously and in different proportions, constituting a normalized vector in a two-dimensional Hilbert space over a complex field C in an orthogonal basis $\{|0\rangle, |1\rangle\}$. According to formula:

$$qubit = |\psi\rangle = A \cdot |0\rangle + B \cdot |1\rangle$$

where $|0\rangle$ and $|1\rangle$ represent the base states: '0' and '1' in ket notation, such that:

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and A, B are the amplitudes of the wave functions for the states: $|0\rangle$ and $|1\rangle$, such that:

$$|A|^2 + |B|^2 = 1.$$

The square of the amplitude A modulus is the probability that the measured qubit will have state $|0\rangle$, and the square of the amplitude B modulus is the probability that the measured qubit will have state $|1\rangle$. The state of such a qubit can also be represented graphically on a Bloch sphere. If we assume that:

$$A = \cos\frac{\theta}{2}, \qquad B = e^{i\varphi}\sin\frac{\theta}{2}$$

then the qubit state can be represented graphically using the normalized state vector $|\psi\rangle$ on the three-dimensional Bloch sphere (Fig. 1).

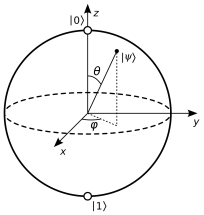


Figure 1. The Bloche sphere (source: [16])

Therefore, the state of the qubit can be uniquely described using two angles: θ and φ. And this form of representing the state of the emulated qubit was adopted by the authors of this article.

4. Model architecture

This chapter will discuss the general architecture of the quantum bit model. In order to test the idea of the model, the entire functional set of the system was designed. It consists of a proper qubit emulator called QUBIT and a device adapted to generate and measure electric digital pulses called GENMET (GENerator-METer).

In addition to these two main elements, the set also includes: a monitor, an external generator, an oscilloscope and a simulation management computer. It is schematically shown in Figure 2. The proposed system emulates the behavior of a virtual qubit in real time and graphically displays the Bloch sphere on the monitor screen along with a vector representing the current state of this qubit. The user has the ability to rotate the qubit's state vector with appropriate input digital pulses, while at the same time the emulated qubit continuously generates its own digital output signals proportional to its current state. Thanks to this, the user has the constant opportunity to observe and read parameters of the qubit, such as its phase and superposition state, which is not directly possible in real quantum systems. Additionally, digital pulses fed into and received from the emulator's outputs can be continuously monitored on the attached oscilloscope.

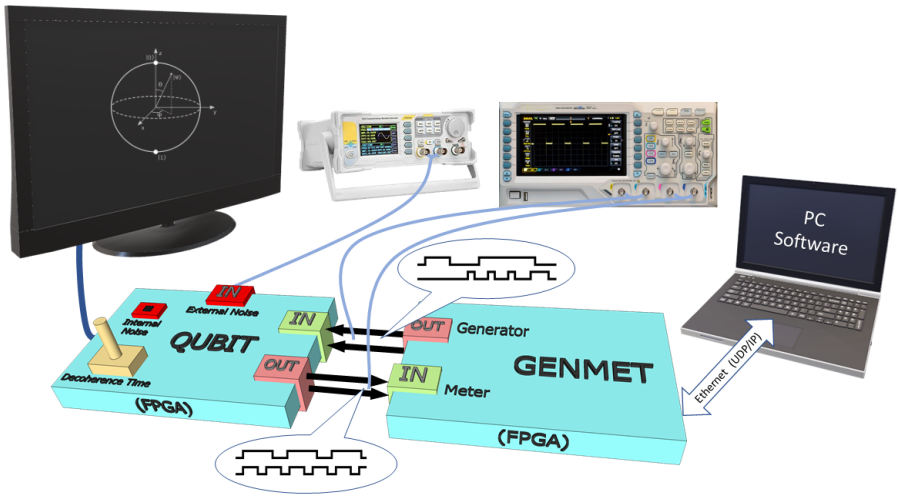


Figure 2. The system idea

QUBIT is also equipped with modules that emulate real physical disturbances. For this purpose, both the internal natural noises of the FPGA system (on the basis of which the emulator was created) are used, as well as special additional circuits that allow the qubit to be disrupted with a user-defined signal generated on an external function generator. Thanks to this, the user can test his quantum operations (algorithms) in the presence of a given interference waveform. It can be, for example, a sinusoidal signal of any frequency, but also white noise or another signal composed of many other harmonic waveforms. In particular, frequency, amplitude or phase modulated signals can be used.

Additionally, to ensure conditions of using the qubit emulator that are close to real ones, the emulator is equipped with a module that controls the qubit coherence time. Thanks to it, the user can set the time after which the qubit will be automatically measured and will lose its quantum state. This situation occurs when working with real quantum systems. For this reason, the user must develop quantum algorithms in such a way that they can be calculated before the system decoheres. The built-in decoherence management module is intended to enable the simulation of this type of situation.

4.1. QUBIT architecture

The QUBIT module, built on the basis of a programmable FPGA integrated circuit, is the main element of the system. It emulates the behavior of a qubit. The state of a qubit is represented by two angles: THETA and PHI. The THETA angle describes the state of superposition of the base states, and the PHI angle determines the phase

of the $|1\rangle$ state in relation to the $|0\rangle$ state. The block diagram of the QUBIT module architecture is presented in Figure 3.

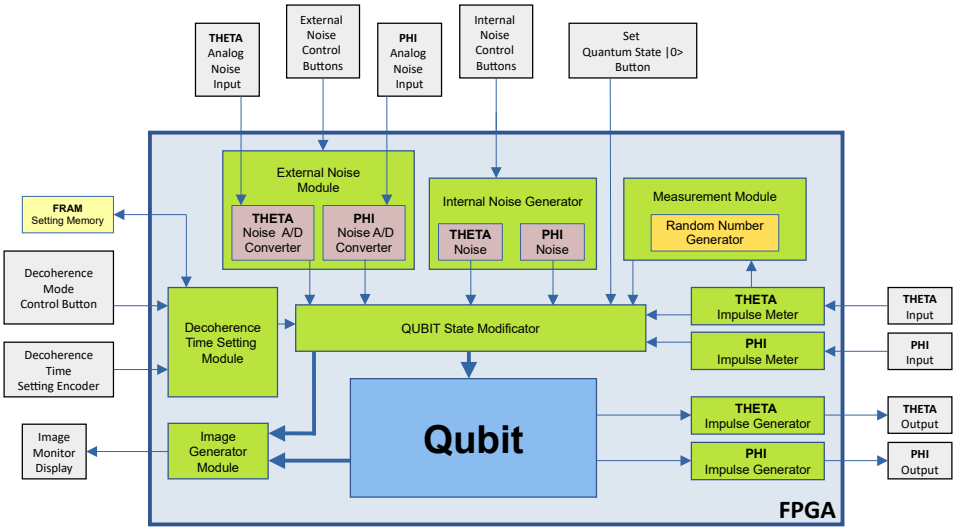


Figure 3. The QUBIT architecture

The QUBIT device is equipped with circuits responsible for handling internal and external noises and with a circuit responsible for emulating the phenomenon of quantum decoherence. The coherence time is set using a dedicated encoder and is stored in the non-volatile FRAM (Ferroelectric Random-Access Memory). The device is also equipped with an image generator that shows the qubit state marked on the Bloch sphere on the monitor screen. The main interaction with the qubit is through the THETA and PHI inputs and outputs.

To reset the THETA angle to zero, i.e. set the qubit to the quantum state $|0\rangle$, a pulse with a length ranging from 9000 µs to 9100 µs (microseconds) must be applied to the THETA digital input.

To hide the qubit state vector in the Bloch sphere, a pulse with a length ranging from 9500 µs to 9600 µs must be applied to the THETA digital input. To restore the visibility of the state vector, a pulse with a length ranging from 9700 µs to 9800 µs must be applied to the THETA digital input.

To rotate the THETA angle by a specific positive angle α, a pulse with a length T [µs] equal to the desired angle expressed in tenths of a degree should be sent to the THETA digital input, i.e. according to the formula:

$$T = \alpha \cdot 10 \text{ [µs]}$$

To rotate the state vector "along" the THETA angle by a specific negative angle α, a pulse of length T [µs] must be sent to the THETA digital input, equal to the

angle α expressed in tenths of a degree and increased by an additional 4000 µs, i.e. according to the formula:

$$T = 4000 + \alpha \cdot 10 \text{ [µs]}$$

For example:

- an angle change of $+43.7^\circ$ requires a pulse length of 437 µs,
- an angle change of -43.7° requires a pulse length of 4437 µs,
- an angle change of -143.7° requires a pulse length of 5437 µs.

To measure the qubit, a pulse with a length between 8000 µs and 8100 µs must be applied to the THETA input. After this operation, the qubit is set to position $|0\rangle$ or position $|1\rangle$, according to the probability represented by the angle θ (THETA).

To zero the PHI angle, i.e. set the qubit phase to 0 degrees (0°), a pulse with a length ranging from 9000 µs to 9100 µs (microseconds) must be applied to the PHI digital input.

To rotate the qubit phase "along" the PHI angle by a specific positive angle α, similarly to the THETA angle, send a pulse of length T [µs] to the PHI digital input, equal to the desired angle expressed in tenths of a degree, i.e. according to the formula:

$$T = \alpha \cdot 10 \text{ [µs]}$$

To rotate the qubit phase "along" the PHI angle by a specific negative angle α, a pulse of length T [µs] must be sent to the PHI digital input, equal to the angle α expressed in tenths of a degree and increased by 4000 µs, i.e. according to the formula:

$$T = 4000 + \alpha \cdot 10 \text{ [µs]}$$

For example:

- an angle change of $+45.1^\circ$ requires a pulse length of 451 µs,
- an angle change of -45.1° requires a pulse length of 4451 µs,
- an angle change of -245.3° requires a pulse length of 6453 µs.

The next ports in the QUBIT module are two digital outputs PHI and THETA, where appropriate electrical impulses are generated continuously and proportionally to the current state of the modeled qubit. The generated pulses for $\theta = 30^\circ$ and $\phi = 45^\circ$ are shown in Figure 4. A non-destructive reading of the current quantum state of the qubit involves measuring the length of the pulses generated at the THETA and PHI outputs. The lengths of these pulses show the values of the corresponding angles, expressed in tenths of a degree, minus one.

Therefore, in order to correctly read the THETA and PHI angles, the pulse lengths (T) measured in microseconds should be reduced by 1 and then divided by 10, i.e. according to the formulas:

$$THETA = \frac{T_{THETA} - 1}{10} \quad [^\circ], \qquad PHI = \frac{T_{PHI} - 1}{10} \quad [^\circ]$$

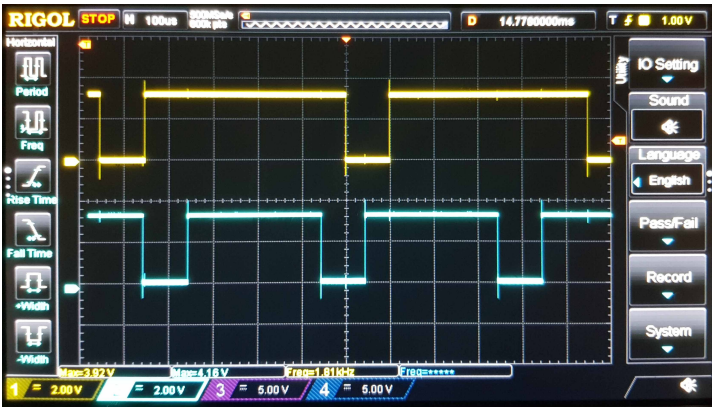


Figure 4. The pulses for $\theta = 30°$ (blue) and $\phi = 45°$ (yellow)

For example:

- The measured pulse of 743 µs indicates an angle of 74.2°
- The measured pulse of 1 µs indicates an angle of 0°

In "Decoherence Mode", the QUBIT module behaves like a real qubit not fully isolated from its surroundings. After switching this mode, information about the currently selected qubit coherence time and information about how much time is left until the system decoheres, appears in the upper right corner of the monitor screen. An example monitor screen in this mode is shown in Figure 5.

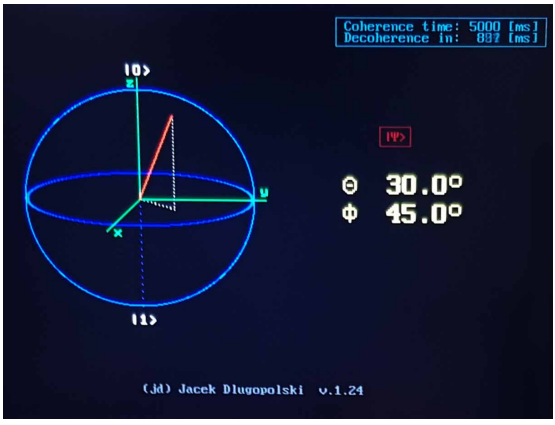


Figure 5. The monitor screen in the decoherence mode

The time during which the qubit is maintained in the quantum state (coherence time) can be changed by the user in the range from 1 ms to 9000 ms with a resolution of one millisecond using a dedicated encoder knob (the current setting is continuously

saved in the non-volatile FRAM memory and is remembered after the power is turned off). This function only works in automatic decoherence mode, i.e. after switching the appropriate switch in the QUBIT module. In this mode, once the qubit has been set to a quantum state, it will automatically decohere after a time set according to the procedure discussed above. This will be equivalent to taking a measurement on it. After decoherence, its classical state will be either "0" or "1", according to the probabilities of the current quantum state just before decoherence. Any operations on the qubit will then no longer work until it is next set to the quantum state.

Another important function of the designed QUBIT is the "Noise" mode, in which the current quantum state of the qubit can be disturbed either by natural noise coming from the internal circuits of the FPGA or by noise artificially generated via an external function generator. Appropriate switches allow you to select the noise source. After turning on the "Noise" mode, a dedicated frame appears on the screen informing about this fact (Fig. 6).

Figure 6. The monitor screen in the noise mode

Then, appropriate switches allow you to turn individual noises on and off, separately for each state component: THETA and PHI.

4.2. GENMET architecture

The QUBIT module discussed in the previous section can be controlled by any pulse generator and read by any pulse length meter. To create a complete system, a dedicated two-channel GENMET device (GENERator-METer) was built, capable of both generating and reading digital electrical impulses. With its help, the full control over the QUBIT module is gain. GENMET communicates with the user via the ETHERNET network, using the UDP/IP protocol. The BROADCAST mode used here makes it unnecessary to set an IP address. Therefore, without any configuration, the system will work in any local subnet. The block diagram of the GENMET module architecture is shown in Figure 7.

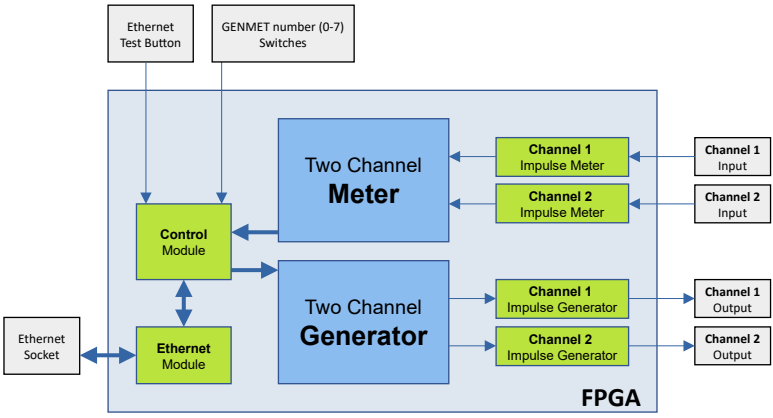


Figure 7. The GENMET architecture

The proposed solution allows for simultaneous operation of up to eight GENMET devices in one local Ethernet subnet. This makes it possible to build circuits consisting of up to eight qubits. GENMET is equipped with an Ethernet connector and four digital ports: two for generating pulses and two for measuring pulses.

GENMET therefore allows for the simultaneous generation of different pulses of a given duration on two independent outputs. The command to generate a pulse with specific parameters can be executed using a PC connected to the module via ETHERNET. The duration (length) of the pulses can be specified in units: from [10 ns] through [µs], [ms] to [s]. After power-on, GENMET is set to microseconds [µs] mode. The outputs should be connected to the appropriate inputs of the QUBIT module and possibly to the appropriate oscilloscope inputs, as schematically shown in Figure 2.

GENMET, regardless of the generation of pulses, also allows for simultaneous two-channel measurement of pulses supplied to two of its inputs. When we connect these inputs to the outputs of the QUBIT module, the GENMET module will continuously measure the length of THETA and PHI pulses coming from the QUBIT module. The values of these measurements can be read at any time using a PC connected via the ETHERNET port.

By connecting a PC to GENMET and using any programming language that supports the UDP protocol, you can immediately control the virtual qubit, read its quantum state and measure it.

GENMET uses the UDP protocol and BROADCAST packets. It expects control commands on port 12000, and sends its responses to port 13000. All commands and responses are sent in text form. It does not require setting an IP address.

GENMET responds to the keyword: **genmet#**, where the "#" sign means a number from 0 to 7. Currently, it is possible to connect up to eight independent

GENMET devices to one local network. The GENMET number "#" is selected using three binary switches on the device.

After sending the "**genmet0 help**" command, we will receive a simplified response informing about the command syntax. From the computer's point of view, we see two channels generating pulses and two channels measuring pulses. We must ensure their proper association (connection) with the THETA and PHI signals of the QUBIT module.

When it comes to the generator of the GENMET device, for each command we can choose the unit in which we will give the duration of the generated pulse. The standard unit is [us] (microseconds). But there are also: [10 ns], [ms] and [s] possible. Selecting [10 ns] means that when specifying a pulse length of 5, we mean a pulse of 50 ns. In addition, we can specify the number of pulses in a single series using the phrase **"nr x"**, where "x" is their number (the default setting is "1", i.e. a single pulse) or instead, we can use the word **"continuous"**, and then the set pulse will be repeated continuously. The pulse parameters on individual channels are defined using the phrases **"low1 x"**, **"high1 x"** and **"low2 x"** and **"high2 x"**. Of course, "1" means the first channel and "2" means the second channel. The "x" variable should be in the range from 0 to 9999 (after power-on, the "x" values are set to 1800 μs). For example, the command:

genmet0 start [us] nr 1 low1 400 high1 650

will generate a single pulse on channel "1" on the GENMET0 device. After sending the command, GENMET will set the digital output of channel "1" to logical zero and wait 400 μs, then set the output to logical one and hold it for 650 μs. After this time, it will set channel "1" to logical zero again.

In order for any impulse to be generated, the word "start" must be included in the command. Missing this word or including the word "show" in the command will ensure that the transmitted command will not cause any changes to the generator outputs. GENMET will then only return information on port 13000 about the current settings. On the other hand, using the word "stop" in the command will stop the currently generated waveform and set the variable "nr" to 0. For example, the command:

genmet0 stop high2

will stop the signal generated on channel "2", and the command:

genmet0 stop all

will stop the generated signals on both channels. In turn command:

genmet0 start [us] nr 5 low 1 400 high1 650

will generate on channel "1" a series of five pulses discussed above. Command:

genmet0 start [us] nr 5 low1 400 high1 650 continuous

will generate 650 µs pulses continuously with 400 µs breaks. This is because when we specify **"nr 5"** and **"continuous"** concurrently, then **"continuous"** has higher priority. When we want to stop generating pulses, we should send the command without **"continuous"**, and it is best to send the command with the phrase **"nr 0"** or with the word **"stop"**. E.g. command:

genmet0 start low1 nr 0 or **genmet0 low1 stop**

To generate a pulse on channel two, you can send the command:

genmet0 start [us] nr 1 low2 541 high2 9000

In this case, one pulse with a duration of 9000 µs, i.e. 9 ms, will appear on the second channel. The last set pulse can be repeated with the command:

genmet0 start high2

It repeats the last set pulse on the second channel. And the command:

genmet0 start high1 high2 or **genmet0 start all**

repeats the last set pulses on both channels. Adding the word "show" will prevent the pulses from being generated, but information about the current settings of the two generator channels will be sent to the PC. Such information about current settings is also always sent when we invoke any valid commands. For example, the command:

genmet0 [us] nr 1 high1 600 low1 400 start

will return the message:

note genmet0 starts [us] nr 1 low1 400 high1 600

informing that a single pulse with a duration of 600 µs was generated on the first channel.

Whereas, command:

genmet0 show [us] nr 1 low1 400 high1 600

will return the message:

note genmet0 shows [us] nr 1 low1 400 high1 600

which informs about the channel settings of the first generator, but the pulse will not be generated.

You can use any UDP terminal or any programming language to communicate with GENMET. Sending a command to GENMET from sample Python code is shown in Figure 8.

```
import socket

s=socket.socket(socket.AF_INET, socket.SOCK_DGRAM)
s.setsockopt(socket.SOL_SOCKET, socket.SO_BROADCAST, 1)

s.sendto(bytes("genmet0 start [us] nr 1 low1 400 high1 2280 ", \
               "utf-8"), ('192.168.0.255', 12000))
```

Figure 8. Sending a command to GENMET in Python

As mentioned earlier, GENMET, in addition to generating pulses, also continuously measures the pulses fed to its two inputs. To read the last measurement, execute the command:

genmet0 read

and then the following text will be sent back to the PC on port 13000:

note genmet0 reads [us] ch1: 682 ch2: 1437

which means that a pulse with a duration of 682 μs recently appeared on the first channel, and a pulse with a duration of 1437 μs on the second channel. If this measurement concerns the QUBIT module, it means that two angles were measured, respectively: 68.1° and 143.6°. The measurement should (according to previous information) be reduced by 1 μs and divided by 10. An example of reading measurement data in Python is shown on Figure 9.

The "data" variable will then contain the information received from GENMET. The presented program prints this information on the computer screen in the last line.

```
import socket

s = socket.socket(socket.AF_INET, socket.SOCK_DGRAM)
s.setsockopt(socket.SOL_SOCKET, socket.SO_BROADCAST, 1)

r = socket.socket(socket.AF_INET, socket.SOCK_DGRAM)
r.setsockopt(socket.SOL_SOCKET, socket.SO_REUSEADDR, 1)
r.setsockopt(socket.SOL_SOCKET, socket.SO_BROADCAST, 1)
r.bind(("",13000))

s.sendto(bytes("genmet0 read ", "utf-8"), ('192.168.0.255', 12000))
data,addr = r.recvfrom(13000)
print(data.decode())
```

Figure 9. Reading measurement data in Python

5. Result

Based on the assumptions and designs of individual elements described above, an actual working prototype of a physical qubit emulation system based on FGPA DE10-Lite [5] modules was built. The effect of this work is shown on Figure 10.

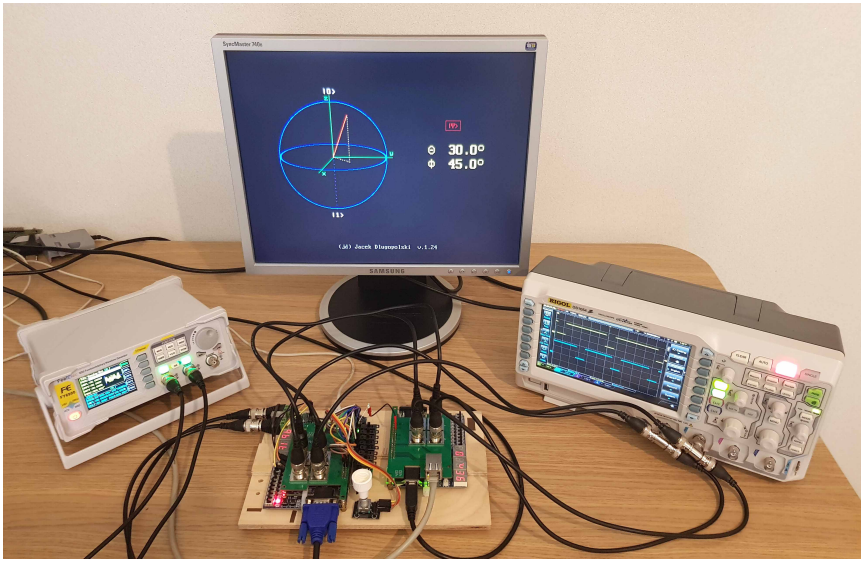


Figure 10. The system prototype

QUBIT and GENMET were built based on two FPGA modules, connected with each other with BNC cables and equipped with devices such as: a monitor, an oscilloscope, a function generator and a PC used to operate the system. The system built in this way was tested and it turned out to work in accordance with the original assumptions. It is possible, among others: setting the qubit coherence time and turning noise on and off, as well as performing any operations on the qubit.

5.1. Dedicated library

Based on the GENMET module control commands described in the previous chapter, the "jd_genmet_lib" library was developed in Python to support the QUBIT module. It provides the basic functions shown in Table 1.

Table 1
The "jd_genmet_lib" library functions

snd(command_string)	Directly send UDP string command to the local ETHERNET network
set_device_name(name)	Set the device name for all of the functions below (default: "genmet0")
read_device_name()	Read the current device name
set_delay_time(t)	Set the delay time between commands in seconds (default is 0.05s)
setup_0_state()	Sets state to: $\|0\rangle$
setup_1_state()	Sets state to: $\|1\rangle$
setup_state(Theta, Phi)	Sets state to: $\cos\left(\frac{Theta}{2}\right) \cdot \|0\rangle + e^{i \cdot Phi} \cdot \sin\left(\frac{Theta}{2}\right) \cdot \|1\rangle$ "Theta" is <0 to 180> degrees and "Phi" is <0 to 360.0> degrees
hide_state()	Hide quantum state
unhide_state()	Unhide quantum state
read()	Read the qubit state (returns the Theta and Phi angles of a state vector)
measure()	Measure the qubit (returns 0 or 1)
degs_to_rad(Theta, Phi)	Convert Theta and Phi from degrees to radians
degs_to_xyz(Theta, Phi)	Convert Theta and Phi from degrees to Cartesian x,y,z
degs_to_amp(Theta, Phi)	Convert Theta and Phi from degrees to apmlitudes of the wave function
rotate_x(Angle)	Rotate X Gate "Angle" is <-360.0 to 360.0> degrees
rotate_y(Angle)	Rotate Y Gate "Angle" is <-360.0 to 360.0> degrees
rotate_z(Angle)	Rotate Z Gate "Angle" is <-360.0 to 360.0> degrees
rotate_theta(Angle):	Rotates Theta between <0 to 180> "Angle" is <-180.0 to 180.0> degrees
rotate_phi(Angle):	Rotate Phi between <0 to 360> "Angle" is <-360.0 to 360.0> degrees
ID()	Identity Gate
X()	Pauli-X (NOT) Gate
Y()	Pauli-Y Gate
Z()	Pauli-Z Gate
XSR()	Square root of Pauli-X Gate –¿ SQRT(NOT)
H()	Hadamard Gate

The user can create new functions and sequences of operations based on this library. For example, after setting the qubit to a specific state, e.g. $\theta = 40°$ and $\phi = 45°$, a non-destructive reading of the quantum state is possible. Ideally, this reading should return the same value on each subsequent read attempt. Such qubit state readings can be made with the "read()" function from the library and, for example, with the simple Python program shown on Figure 11.

```
import time
import datetime
import string
import math
from jd_genmet_lib import *

def do_log(string1):
  t = datetime.datetime.fromtimestamp(time.time()).strftime('%H:%M:%S:%f')
  log = open(".\\data.txt", "a")
  print(t, " -> ", string1, file=log);log.flush()
  print(t, " -> ", string1)

log = open(".\\data.txt", "a"); print("\n\n", file=log);log.flush()
set_device_name("genmet0")
str1="   Device used:" + str(read_device_name()) + "\n"; do_log(str1)

str1=("   Theta   Phi");  do_log(str1)
for i in range(0,10):
  r=read() # Read the qubit
  Theta,Phi=r;
  str1=("   %5.1f   %5.1f" % (Theta,Phi) );do_log(str1)
  time.sleep(0.5)
```

Figure 11. Qubit state reading in Python

Figure 12a shows the result of sample readings without noise enabled. After turning on qubit internal noise for both parameters: θ and ϕ, subsequent readings will be distored, as shown in Figure 12b.

a)

```
13:28:27:435917  ->     Device used:genmet0

13:28:27:435917  ->     Theta   Phi
13:28:27:503460  ->      40.0    45.0
13:28:28:055694  ->      40.0    45.0
13:28:28:607882  ->      40.0    45.0
13:28:29:160037  ->      40.0    45.0
13:28:29:712618  ->      40.0    45.0
13:28:30:264841  ->      40.0    45.0
13:28:30:817210  ->      40.0    45.0
13:28:31:369679  ->      40.0    45.0
13:28:31:921846  ->      40.0    45.0
13:28:32:474403  ->      40.0    45.0
```

b)

```
13:35:28:347785  ->     Device used:genmet0

13:35:28:347785  ->     Theta   Phi
13:35:28:412296  ->      39.8    44.9
13:35:28:963894  ->      39.8    44.8
13:35:29:516330  ->      39.9    44.8
13:35:30:068559  ->      39.6    44.9
13:35:30:621107  ->      39.8    44.8
13:35:31:173512  ->      39.8    45.2
13:35:31:726022  ->      39.9    44.9
13:35:32:278265  ->      39.9    44.8
13:35:32:830630  ->      40.1    44.9
13:35:33:383059  ->      39.8    44.6
```

Figure 12. Sample readings without noise (a); sample reading with internal noise (b)

The library also allows the conversion of parameters: θ and ϕ into the amplitudes of the base states $|0\rangle$ and $|1\rangle$. The function: "degs_to_amp(Theta, Phi)" can be used for this purpose. An example of using this function is shown in Figure 13. The cnversion results are shown in Figure 14. Readings were also made here with internal noise turned on. Of course, the qubit state disturbed in this way will also affect the quantum measurement made using the "measure()" function available in the described library.

```
import time
import datetime
import string
import math
from jd_genmet_lib import *

def do_log(string1):
  t = datetime.datetime.fromtimestamp(time.time()).strftime('%H:%M:%S:%f')
  log = open(".\\data.txt", "a")
  print(t, " -> ", string1, file=log);log.flush()
  print(t, " -> ", string1)

log = open(".\\data.txt", "a"); print("\n\n", file=log);log.flush()
set_device_name("genmet0")
str1="   Device used:" + str(read_device_name()) + "\n";do_log(str1)

str1=("   Theta   Phi        |0>            |1>");do_log(str1)
for i in range(0,10):
  r=read() # Read the qubit
  Theta,Phi=r; ar,br,bi=degs_to_amp(Theta,Phi)
  str1=("   %5.1f   %5.1f     %13.10f  %13.10f %+13.10f *i" % \
       (Theta,Phi,ar,br,bi) );do_log(str1)
  time.sleep(0.5)
```

Figure 13. Converting read angles to amplitudes in Python

```
13:44:16:285992  ->     Device used:genmet0

13:44:16:285992  ->     Theta   Phi        |0>            |1>
13:44:16:340292  ->      39.8    44.8      0.9402881270   0.2415233681 +0.2398430755 *i
13:44:16:891826  ->      40.1    44.8      0.9393937941   0.2432692664 +0.2415768274 *i
13:44:17:444393  ->      39.9    45.0      0.9399907318   0.2412648173 +0.2412648173 *i
13:44:17:996964  ->      39.9    44.9      0.9399907318   0.2416855361 +0.2408433635 *i
13:44:18:549387  ->      39.8    44.9      0.9402881270   0.2411043954 +0.2402642477 *i
13:44:19:101808  ->      39.8    44.9      0.9402881270   0.2411043954 +0.2402642477 *i
13:44:19:654333  ->      39.9    44.9      0.9399907318   0.2416855361 +0.2408433635 *i
13:44:20:206770  ->      39.8    44.9      0.9402881270   0.2411043954 +0.2402642477 *i
13:44:20:759131  ->      39.8    44.9      0.9402881270   0.2411043954 +0.2402642477 *i
13:44:21:311241  ->      39.8    44.8      0.9402881270   0.2415233681 +0.2398430755 *i
```

Figure 14. Printout of converted amplitudes

5.2. Application examples

The main goal of the developed and built QUBIT was to create a research, educational and training environment for carrying out a number of experiments allowing future users of real quantum systems to become familiar with potential problems occurring in real quantum systems. Additionally, having such a device that does not require specialized low-temperature laboratory conditions makes it easier to look for algorithms and solutions to counteract the problems mentioned above.

A single QUBIT can, for example, be used to practice generating appropriate control pulses that allow the qubit state vector to rotate around any selected axis on the Bloch sphere. This will allow you to independently develop any single-qubit gates.

QUBIT is equipped with advanced physical interference features, discussed earlier. Hence, another example of using a single QUBIT is to use it for developing effective noise reduction algorithms in quantum systems. Finding an effective solution for a single qubit here can then scale to multi-qubit systems.

In addition, the device has been designed to also enable the creation of multi-qubit environments. GENMET, discussed in the previous chapters, has three special switches to set its unique name under which it is seen in the local Ethernet network. Thanks to this, up to eight independent GENMET devices can now be connected to one network (with minor modifications, this number can be significantly increased). Then each GENMET can control a separate QUBIT. This is schematically shown in Figure 15.

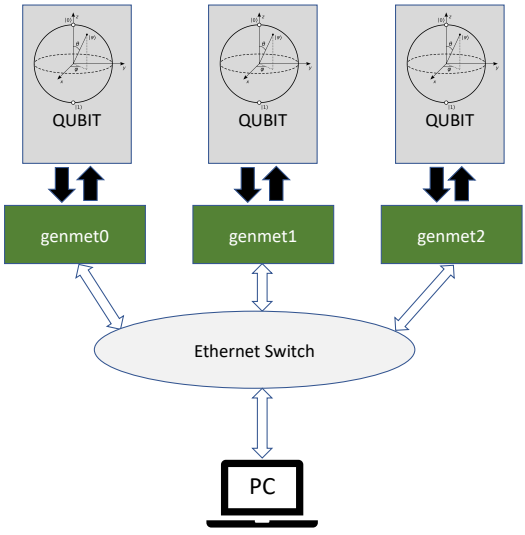


Figure 15. The multiQUBIT application

With such a configuration, an additional layer integrating all connected QUBITs can be created on any connected PC and based on any programming language This will make it possible to emulate entanglement between QUBITs and study various multi-qubit algorithms. Work of this type is currently underway, and the results will be presented in future articles.

6. Conclusion

To sum up, the constructed qubit model is much closer to the real quantum system than any other purely IT model. The change in the state of the qubit does not occur by simply sending selected values to the model, but requires the application of electrical impulses with precisely defined physical parameters, in particular the exact duration of these pulses. So this is exactly as is currently required in real quantum systems. The same applies to reading the state of a qubit. Such a readout requires a physical measurement of the duration of the pulses continuously generated by the qubit. During the process of forming and transmitting impulses, such a model is exposed to all kinds of disturbances and difficulties encountered in all kinds of real physical systems. No quantum system simulators that we are currently aware of offer this possibility.

Additionally, the presented model is equipped with mechanisms for intentionally generating noise in the system, both based on the internal noise of real qubits and based on noise usually coming from the environment. Natural noise occurring inside the FPGA circuitry is used to emulate the internal noise of the qubit. It is thanks to the naturalness of the phenomena occurring there that the process of generating internal noise is characterized by actual physical randomness.

On the other hand, any external electrical waveform generator is used to simulate ambient noise, so by selecting the appropriate waveform or combining a total noise signal from many different harmonics, many real-world phenomena that could potentially interfere with quantum systems can be physically simulated.

An additional advantage of the presented model is the ability to impose a coherence time regime, after which the model automatically loses its quantum state by making spontaneous self-measurements. This necessitates the real need to limit the number of quantum operations during one experiment session.

In addition, the ability to connect up to eight such qubits to a single local network simultaneously (and this number can be easily significantly expanded) allows the quantum circuits emulated in this way to be expanded to much more complex multi-qubit systems.

All the elements discussed above mean that the described model can be successfully used for initial emulation and study of quantum phenomena, before it is possible to safely start real experiments on real quantum systems.

References

[1] Aminian M., Saeedi M., Saheb Zamani M., Sedighi M.: FPGA-Based Circuit Model Emulation of Quantum Algorithms. In: *2008 IEEE Computer Society Annual Symposium on VLSI*, pp. 399–404, 2008. doi: 10.1109/ISVLSI.2008.43.

[2] Burgholzer L., Bauer H., Wille R.: Hybrid Schrödinger-Feynman Simulation of Quantum Circuits With Decision Diagrams. In: *2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*, IEEE, 2021. doi: 10.1109/qce52317.2021.00037.

[3] Chapeau-Blondeau F.: Modeling and Simulation of a Quantum Thermal Noise on the Qubit, *Fluctuation and Noise Letters*, vol. 21(06), 2022. doi: 10.1142/s0219477522500602.

[4] Cheng S., Cao C., Zhang C., Liu Y., Hou S.Y., Xu P., Zeng B.: Simulating noisy quantum circuits with matrix product density operators, *Physical Review Research*, vol. 3(2), 2021. doi: 10.1103/physrevresearch.3.023005.

[5] DE10-Lite Board. https://www.terasic.com.tw/cgi-bin/page/archive.pl?Language=English&CategoryNo=218&No=1021. Accessed: 14.4.2024.

[6] Fujishima M., Inai K., Kitasho T., Hoh K.: 75- qubit Quantum Computing Emulator. In: *Extended Abstracts of the 2003 International Conference on Solid State Devices and Materials*, pp. 406–407, 2003. doi: 10.7567/SSDM.2003.P1-4.

[7] Grurl T., Fuß J., Hillmich S., Burgholzer L., Wille R.: Arrays vs. Decision Diagrams: A Case Study on Quantum Circuit Simulators. In: *2020 IEEE 50th International Symposium on Multiple-Valued Logic (ISMVL)*, pp. 176–181, 2020. doi: 10.1109/ISMVL49045.2020.000-9.

[8] Hillmich S., Markov I.L., Wille R.: Just Like the Real Thing: Fast Weak Simulation of Quantum Computation. In: *2020 57th ACM/IEEE Design Automation Conference (DAC)*, IEEE, 2020. doi: 10.1109/dac18072.2020.9218555.

[9] IBM Qiskit Noise Models, https://qiskit.github.io/qiskit-aer/apidocs/aer_noise.html#quantum-error-functions. Accessed: 3.07.2024.

[10] IBM Quantum Learning – Composer, https://quantum-computing.ibm.com/composer/files/new. Accessed: 14.4.2024.

[11] Negovetic G., Perkowski M., Lukac M., Buller A.: Evolving Quantum Circuits and an FPGA-based Quantum Computing Emulator. 2002. https://api.semanticscholar.org/CorpusID:15995633.

[12] Quantum AI team and collaborators: qsim, 2020. doi: 10.5281/zenodo.4023103.

[13] Quirk: Quantum Circuit Simulator, https://algassert.com/quirk. Accessed: 14.4.2024.

[14] The Quantum Länd: Quantum Circuit Simulator, https://thequantumlaend.de/frontend/designer.php. Accessed: 14.4.2024.

[15] Wei K., Niwase R., Amano H., Yamaguchi Y., Miyoshi T.: A state vector quantum simulator working on FPGAs with extensible SATA storage. In: *2023 International Conference on Field Programmable Technology (ICFPT)*, pp. 272–273, 2023. doi: 10.1109/ICFPT59805.2023.00041.

[16] Wikipedia: Bloch sphere. https://en.wikipedia.org/wiki/Bloch_sphere.
[17] Xie T., Zhao Z., Xu S., Kong X., Yang Z., Wang M., Wang Y., Shi F., Du J.: 99.92%-Fidelity CNOT Gates in Solids by Noise Filtering, *Physical Review Letters*, vol. 130, 030601, 2023. doi: 10.1103/PhysRevLett.130.030601.
[18] Zulehner A., Wille R.: Advanced Simulation of Quantum Computations, *arXiv*, 2018. doi: 10.48550/arXiv.1707.00865.

Affiliations

Jacek Długopolski
AGH University of Krakow, Faculty of Computer Science, al. A. Mickiewicza 30, 30-059 Krakow, Poland, dlugopol@agh.edu.pl

Marcin Sadowski
SONOVERO R&D, www.sonovero-rd.pl, Warsaw, Poland, msadowski@sonovero-rd.pl

Wawrzyniec Suleja
SONOVERO R&D, www.sonovero-rd.pl, Warsaw, Poland, wsuleja@sonovero-rd.pl

Received: 6.05.2024
Revised: 29.07.2024
Accepted: 29.07.2024

Computer Science • 25(4) 2024 https://doi.org/10.7494/csci.2024.25.4.6070

Nyi Nyein Aung
Wanus Srimaharaj

DISTANCE-BASED INTEGRATION OF ERP CORRELATION ANALYSIS

Abstract *Interpretation of cognitive performance is a paramount pursuit in learning achievements. Cognitive abilities, encompassing attention, memory, decision-making, and language comprehension, are recognized on individual's capacity to navigate in diverse cognitive tasks. In the academic domain, optimal cognitive functioning is essential for effective learning, information retention, and problem-solving. Proficiency in cognitive skills is directly linked to academic success and intellectual development, providing the necessary cognitive tools for processing, and synthesizing complex information. Therefore, this study explores the correlation between event-related potential (ERP) sub-components (P300, N170, N400) to assess the intricacies of cognitive performance. A regularized approach utilizing Spearman's Rank Correlation Coefficient and Euclidean Distance is employed. Positive correlations reveal consistent relationships among P300, N170, and N400 ranks across ERP waveforms, indicating similar response patterns. Negative correlations denote inverse relationships. Moreover, the theoretical framework focuses on the digital filtering, ensemble averaging, and baseline correction from data contrast discrimination tasks. Findings indicate positive correlations, suggesting higher ERP amplitudes correspond to superior cognitive performance. This tailored and integrated methodology, indicating the correlation between ERP sub-components, contributes to the broader field of neuroscience and informatics, potentially informing cognitive enhancement strategies in education and bio-medical analysis.*

Keywords Spearman's rank correlation coefficient, correlation analysis, Euclidean distance, cognitive performance, event-related potentials, ERP, P300, N170, N400

Citation Computer Science 25(4) 2024: 521–545

1. Introduction

Cognitive performance, encompassing processes such as attention, memory, perception, and decision-making, is a fundamental aspect of human functioning. The brain's ability to process and interpret information is vital for problem-solving, learning, and decision-making, impacting academic and professional success, daily functioning, and overall quality of life. High cognitive performance facilitates efficient information processing, effective decision-making, rapid learning, and adaptation to changing environments. Conversely, cognitive impairment can hinder independence, social interactions, and daily tasks, particularly as individuals age. Research in assessing and enhancing cognitive performance has grown in importance, aiming to optimize brain function and promote cognitive well-being.

Event-Related Potentials (ERPs) are specific patterns of brain activity measured through electroencephalography (EEG) in response to sensory, cognitive, or motor events. These brain responses are time-locked to stimuli, allowing researchers to examine the timing and stages of cognitive processing. ERPs are characterized by distinct wave components such as P1, N1, P2, N2, P3 (P300), N400, and LPC, each associated with different cognitive functions like sensory processing, attention, memory, and language comprehension. The amplitude and latency of these components provide insights into the intensity and speed of cognitive processes. ERPs are essential for understanding cognitive performance as they reflect the brain's real-time processing of information. Variations in P300 amplitude and latency are key indicators of learning performance across tasks involving visual or auditory discrimination, semantic processing, and spatial navigation. Changes in coherence patterns reflect alterations in neural synchronization and information transfer, key elements in learning and memory processes.

The P300 is a positive peak in the ERP waveform occurring approximately 300 milliseconds after stimulus presentation, associated with cognitive processes like attention, memory, and decision-making. Detecting the P300 typically involves measuring brain activity during experiments where subjects are exposed to various stimuli, often using the oddball paradigm. P300 is widely studied concerning learning performance and is associated with working memory capacity and attentional control.

The N170 is an ERP sub-component elicited by faces and other visual stimuli, essential in face processing and social cognition. Its amplitude and latency can be influenced by factors like familiarity, attention, and expertise, indicating its sensitivity to changes in perceptual and cognitive processing during learning. Similarly, the N400, elicited by auditory stimuli, is associated with early auditory processing, including sound localization and discrimination. The N400's amplitude and latency can be modulated by factors such as stimulus complexity, attention, and expertise, suggesting its sensitivity to changes in auditory processing during learning. The N170 amplitude is linked to face recognition ability, and the N400 amplitude relates to semantic processing and language comprehension.

The correlations between these ERP sub-components and learning performance offer potential biomarkers for assessing cognitive abilities in educational and training contexts. By measuring the amplitudes and latencies of these ERP sub-components, researchers can investigate the cognitive processes involved in different tasks and stimuli. Clinically, ERPs are used to diagnose and monitor neurological and psychiatric conditions, providing valuable data on the effectiveness of cognitive training and rehabilitation programs. ERP sub-components like P300, N170, and N400 serve as valuable biomarkers for cognitive performance assessment and enable a deeper exploration of the neural mechanisms underpinning cognitive processes. Analyzing ERP components provides detailed insights into neural mechanisms underlying cognitive functions, making ERPs a powerful tool in both research and clinical contexts.

This study aims to investigate the correlations between P300, N170, and N400, which are ERPs associated with various aspects of cognitive performance. By analyzing the amplitudes and latencies of these ERP sub-components, the study seeks to provide insights into the neural mechanisms underlying cognitive processes related to attention, memory, perception, and decision-making. The goal is to utilize these ERP sub-components as biomarkers to assess cognitive abilities and enhance educational and training strategies, contributing to a better understanding of cognitive performance and its potential applications in optimizing brain function and cognitive well-being.

2. Literature review

Identifying human cognitive performance can be done through various methods, including classification by general process involved, regional brain functions, and hierarchical structure based on the complexity of operations. ERP measures the brain's response to specific stimuli [17, 18]. This allows investigators to explore a nearly infinite number of domains where it is of interest to understand the relative timing of neural events in a non-invasive method [11].

In ERP, the stimulus is presented multiple times to the participant, and the responses are measured. ERPs are created by averaging responses to standard and deviant stimuli separately. ERPs measure voltage changes in the brain that occur following the onset of specific stimuli or cues and provide a measure of the timing of the brain's communication or timing of information processing. It is extensively used in neuroscience, cognitive psychology, cognitive science, and psychophysiological research to measure cognitive performance. The amplitude, latency, and topography of the resulting positive and negative deflections are taken to index the underlying mental operations. ERPs provide a continuous measure of processing between a stimulus and a response, allowing us to determine which stage(s) are being affected by a specific experimental manipulation.

Additionally, ERP records brain processes on a millisecond scale, capturing neural activity related to both sensory and cognitive processes. It is used in experimental

settings and is involved in language research [36]. ERP can be associated with sensory encoding, inhibitory responses, updating working memory, and highlighting the temporal unfolding of neural activity associated with different cognitive aspects of language comprehension and production [23]. Moreover, studies have estimated the test-retest reliability of ERP waves, with interclass correlation coefficients between first and second recordings being around 0.8 for amplitude and around 0.9 for the latency of the P3 NOGO waves, indicating the reliability of ERPs as measures of brain functioning [6, 24].

Classification by general process involves memory, attention, language, and executive functioning, while regional brain functions are derived from lesion studies and include the frontal lobe, temporal lobe, parietal lobe, hippocampus, or other structures [20, 21]. Cognitive ability domains can also be conceptualized in several ways, such as a hierarchical structure based on the complexity of operations, with basic sensory and perceptional operations being the least complex and reasoning and problem-solving being the most complex [14]. Tests of general cognitive ability are used to identify human cognitive performance. The most used cognitive tests usually take 15 minutes or less and include repeating lists of words or spelling words backward [31]. These tests are good predictors of job performance and training success for a wide variety of jobs. Processing speed is the strongest predictor of overall cognitive performance and is correlated with impairments in everyday functioning, with coding tasks showing the most significant impairment in schizophrenia. However, there are inconsistencies in the clinical and research literature, especially in broad domains that may include multiple component processes [39]. In addition, there is an issue with the intrinsic validity of cognitive domains in populations other than those with specific regional brain damage [4]. Contemporary circuit-based conceptions focus on the activation and interaction of these circuits. Current methods such as smartphone assessment and remote cognitive assessment are more convenient for longitudinal assessment and can measure preclinical AD-related changes in long-term associative memory across varied memory retention intervals [35].

Moreover, there are a variety of methods available for measuring cognitive performance, including gamified assessment, smartphone assessment, and assessments of GPS data and gait characteristics [35]. Smartphone assessment is useful in measuring preclinical AD-related changes in long-term associative memory but requires retention intervals of at least 3 days to be sufficiently sensitive to differences in recall and recognition performance in adults without diagnosed cognitive impairment. Gamified assessment has been found to reduce testing anxiety and increase task engagement and enjoyment without affecting performance and can provide better construct and ecological validity than simple laboratory-based tasks thanks to a more realistic context [26]. Assessment of GPS data and gait characteristics measured through wearable accelerometers have also been found to differentiate among dementia subtypes with moderate accuracy, while recent developments allow neuropathology associated with potential cognitive decline to be accurately detected from peripheral

blood samples [28]. However, the accuracy of these methods can be affected by various factors such as device type, operating system, and Wi-Fi connection, as well as subtle differences in task design and the lack of interoperability between cognitive functioning metrics.

In conclusion, while there are a variety of methods available for measuring cognitive performance, researchers need to carefully consider the strengths and limitations of each method to accurately assess cognitive function. The application of correlation and distance analysis holds great promise in advancing our understanding of cognitive processes associated with various tasks. By adopting these methods, this study can gain valuable insights into the limitations of traditional techniques such as ERP. The existing methods highlighted the strengths and limitations of existing methods, emphasizing the need for alternative approaches to studying cognitive performance. In this study, Spearman's rank correlation coefficient offers a statistical measure to examine the relationships between cognitive performance measures and other variables, providing valuable insights into cognitive processes. On the other hand, Euclidean distance analysis enables the assessment of similarity or dissimilarity in cognitive profiles, paving the way for future research and the development of diagnostic tools for cognitive impairments. By leveraging these methods, this study can further our understanding of cognitive processes, improve diagnostic accuracy, and enhance clinical interventions for individuals with cognitive impairments.

In summary, the research aims to employ correlation and distance analysis techniques to gain a deeper understanding of cognitive performance, bypassing the limitations of traditional methods like ERP. The study aims to explore relationships between cognitive performance measures and other variables using Spearman's rank correlation coefficient. Additionally, it seeks to assess the similarity or dissimilarity in cognitive profiles through Euclidean Distance analysis. Nevertheless, these methods can contribute to a more comprehensive understanding of cognitive processes and have the potential to enhance diagnostic accuracy and clinical interventions for individuals with cognitive impairments.

3. Theoretical framework

3.1. Cognitive performance and related ERP sub-components

Cognitive performance denotes an individual's capacity to process and utilize information, encompassing a spectrum of skills, including attention, memory, decision-making, problem-solving, and language comprehension effectively and efficiently. The evaluation of cognitive performance involves a diverse array of methods, spanning behavioral assessments, neuroimaging modalities, and electrophysiological measures, event-related potentials (ERPs) [37]. ERPs provide insights into the brain's electrical activity concerning specific events or stimuli. Among these ERPs, the P300, N170, and N400 stand as prominent sub-components frequently employed in the investigation of cognitive performance.

The P300, or P3 wave, manifests as a positive ERP sub-component, surfacing approximately 300 ms post-presentation of a target stimulus [29]. It is conventionally associated with cognitive processes such as attention, working memory, and decision-making. The amplitude and latency of the P300 serve as indicators of cognitive performance, with larger and faster P300 responses signifying enhanced cognitive processing and performance.

In contrast, the N170, a negative ERP sub-component, emerges approximately 170 ms following the introduction of a visual stimulus, typically a facial image [16]. Its primary function lies in the processing of facial features and the recognition of faces. The amplitude and latency of the N170 serve as markers of cognitive performance, with larger and more rapid N170 responses signaling improved face processing and recognition.

Conversely, the N400, another negative ERP sub-component, materializes around 400 ms after the presentation of a semantic stimulus, such as a word or sentence [34]. The N400 is intimately connected with semantic processing, encapsulating language comprehension, and memory retrieval. Here again, the amplitude and latency of the N400 offer insights into cognitive performance, with larger and more rapid N400 responses indicative of superior semantic processing and comprehension. Each ERP waveform boasts unique characteristics and is influenced by specific factors, as outlined in Table 1.

Table 1

Main characteristics of P300, N170, and N400 waveforms [19, 23, 30]

ERP waveforms	Latency [ms]	Amplitude range [μV]	Characteristics and factors
P300	250–350	5–25	Reflects attention, cognitive processing, and task relevance. Amplitude can vary based on stimulus characteristics, attentional demands, and individual differences
N170	120–200	2–10	Typically observed in response to visual stimuli, particularly faces. Amplitude can be influenced by facial familiarity, emotional expression, and attentional focus
N400	300–500	2–8	Occurs in response to semantically meaningful stimuli. Amplitude is affected by semantic processing, stimulus congruity, and contextual integration

Table 1 offers amplitude ranges as general reference points, drawing from typical observations in prior [19, 23, 30]. However, it is essential to acknowledge that actual amplitude values may exhibit variation contingent upon the specific experimental

protocols, recording configurations, and individual dissimilarities. These amplitude ranges, therefore, serve as an initial framework for comprehending these ERPs and the multifaceted determinants affecting the magnitudes. The P300, N170, and N400 ERP sub-components stand as invaluable instruments in elucidating the cognitive processes and performance of individuals within cognitive neuroscience and clinical investigations. Figure 1 shows a simulated ERP waveform with the N170, P300 and N400 sub-components.

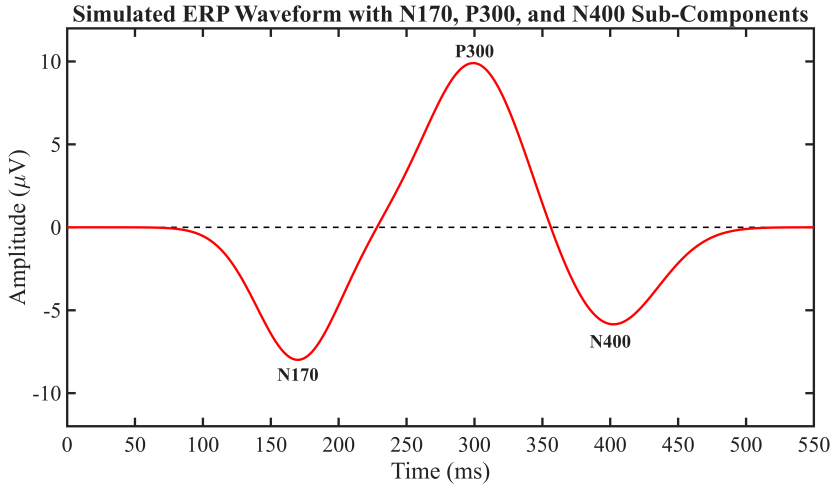


Figure 1. Simulated ERP Waveform with N170, P300 and N400 Sub-Components

3.2. Correlation coefficient

Spearman's rank correlation coefficient is a statistical measure assessing the strength of the relationship between two variables [15]. In the context of human cognitive performance and ERP, it determines the association between the amplitude or latency of ERP sub-components (e.g., P300, N170, N400) and measures of cognitive performance, such as reaction time or accuracy. Spearman's rank correlation coefficient relies on rank order rather than specific numerical values. The coefficient ranges from -1 (perfect negative correlation) to $+1$ (perfect positive correlation), with 0 indicating no correlation. It also reveals the proportion of variability in one variable explained by another. In a study measuring P300 component amplitude in an attention-demanding task and participants' reaction times to target stimuli, Spearman's rank correlation coefficient assesses the association between P300 amplitude and reaction time. A positive coefficient suggests that as P300 amplitude increases, reaction time decreases, signifying better cognitive performance. Therefore, Spearman's rank correlation coefficient is a valuable statistical tool for examining the relationship between ERP sub-components and cognitive performance in studies exploring the neural basis of cognitive processes.

3.3. Distance-based measurement

Euclidean Distance (ED) is a widely employed method in numerous academic disciplines, renowned for its versatility in quantifying the similarity or dissimilarity between data points [27]. Within the domain of cognitive neuroscience, particularly in the analysis of Event-Related Potentials (ERPs) and cognitive performance, ED emerges as a crucial metric [7]. This distance metric, computed by measuring the straight-line separation between two points in a multidimensional space, allows for a nuanced comparison of ERP waveforms, where each data point signifies the waveform's amplitude at a specific time point [7]. The application of ED in the analysis of cognitive performance and ERP sub-components is supported by various empirical studies. For instance, [22] demonstrate the effectiveness of different distance metrics, including Euclidean, in comparing ERP waveforms, while [30] provides insights into the relationship between ERP components and cognitive processes. Smaller ED between ERP waveforms obtained from individuals engaged in cognitive tasks often indicates greater similarity in cognitive performance [7]. Conversely, larger distances suggest greater dissimilarity, implying differences in cognitive processing. Moreover, changes in ED pre and post-cognitive interventions offer valuable insights into the impact of these interventions on cognitive performance. For instance, studies by [12] explore the effects of cognitive interventions on ERP components, shedding light on their potential to modulate cognitive processes. Through the meticulous analysis of ERP waveforms using ED, researchers can unravel the intricate neural mechanisms underlying cognitive processes and their susceptibility to various modulations, including cognitive interventions and individual differences [7]. Thus, ED stands as an indispensable tool in exploring the complex relationship between cognitive performance and ERP sub-components, supported by empirical evidence from cognitive neuroscience research.

4. Methodology

The research framework, as illustrated in Figure 2, comprises three primary stages: 1) data collection, 2) data preprocessing, and 3) data analysis. During the data preprocessing phase, digital filtering, ensemble averaging techniques, and baseline correction are implemented to extract the ERP sub-components from the raw data. Subsequently, in the data analysis stage, the Correlation and Distance-Based approach is employed on the preprocessed ERP waveforms, and the resultant findings are subjected to rigorous statistical testing for significance.

4.1. Data collection

In conventional psychological studies, data collection typically necessitates the involvement of expert psychologists to ensure adherence to ethical standards. However, this study adopted an open-source event-related potential (ERP) dataset obtained through contrast discrimination tasks [1].

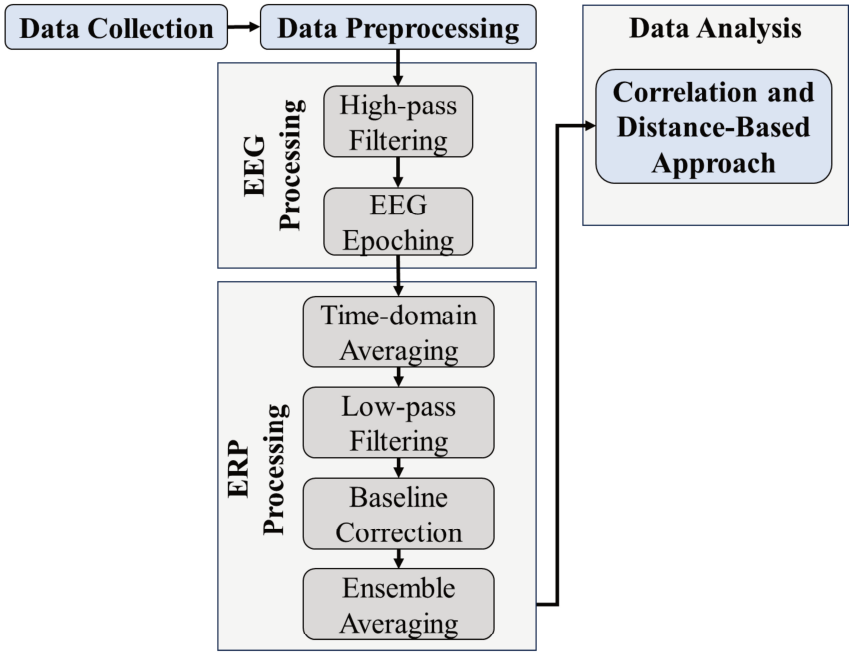


Figure 2. Framework for ERP Analysis Using Correlation and Distance-Based Approach

This innovative approach allowed the investigation to proceed without the direct involvement of specialized psychological experts. The raw brain signal data was acquired using a WaveGuard cap and an ANT Neuroscan EEG system, which incorporated 64 electrodes. While the 10/20 system generally refers to a method for electrode placement, it does not inherently consist of 64 electrodes. Instead, the electrodes were arranged in a manner consistent with the 10/20 system principles, ensuring comprehensive coverage of the scalp. The AFz electrode served as the ground reference. Data acquisition was performed at a sampling rate of 1 kHz utilizing the ASALab software. To pinpoint the P300, N170, and N400 components, electrode placements following the 10/20 system were adopted, covering prefrontal (Fp), frontal (F), temporal (T), parietal (P), occipital (O), and central (C) regions. Additionally, midline sagittal plane electrodes denoted as 'Z' (zero), including FpZ, Fz, Cz, and Oz, were designated as grounding or referencing points. These electrode selections were meticulously made to encompass an extensive array of brain regions relevant to the P300, N170, and N400 components. Channels Pz, P3, P4, and Cz were specifically chosen to effectively capture neural activity linked to the P300 component, known to manifest in these regions [8]. The N170 component, primarily associated with face processing, was targeted using channels Oz, O1, O2, P7, and P8, selected for their significance in the occipital and posterior temporal regions [2]. For the detection of the N400 component, channels Cz and CPz were included, as these central and centroparietal

regions are commonly associated with substantial N400 activity [23]. During the experimental procedure, participants were seated in a dimly lit room 57 cm from the display screen. Responses were collected through a mouse interface to indicate observations. The task was divided into five blocks, each comprising an 8-minute trial period with 200 trials per target contrast. Trigger events with infrequent occurrences during human brain signal detection were excluded from the analysis in this study. To summarize, meticulous selection of electrode placements and the experimental design were devised to concurrently capture the P300, N170, and N400 components. These specific channels were chosen considering the distinct requirements for identifying and investigating these components, all within the broader framework of the electrode placement system employed in this study. It is important to note that the experiment excluded the last 25% of the trials (50 trials) due to unusually high noise levels caused by external factors.

4.2. Data preprocessng

4.2.1. EEG processing

In the raw brain signal recordings, a significant presence of noise is a common occurrence, primarily stemming from participant movements and voltage dispersion between different electrodes during data collection. This noise can be effectively eliminated from the raw Brain Signals to enhance the extraction of ERP sub-components. One approach employed for noise reduction is the use of finite impulse response (FIR) filters, which are digital filters capable of removing noise that contaminates the raw Brain Signals across a wide frequency range. FIR filters exhibit linearity in the phase, ensuring enhanced stability during the filtering process, a feature that distinguishes them from infinite impulse response (IIR) filters. Typically, a high-pass filter with a cutoff frequency of 0.1 Hz and a low-pass filter with a cutoff frequency of 30 Hz is applied to brain signal recordings to attain an optimal noise-to-signal ratio. These filters effectively eliminate signals with frequencies falling below or exceeding the specified cutoff frequencies, optimizing the overall signal quality. The convolution equation representing the operation of a finite impulse response filter is presented in Equation (1). This mathematical representation illustrates the process by which FIR filters contribute to the refinement of brain signal data, enhancing the extraction of ERP sub-components.

$$y(n) = \sum_{k=0}^{N-1} \mathrm{b}_k \cdot x(n-k) \tag{1}$$

The elements of the equation can be defined as follows: $y(n)$ represents the output signal, $x(n)$ corresponds to the input signal, N signifies the filter order, and b_k denotes the value of the impulse response at the instance denoted as k. These components collectively illustrate the fundamental parameters and variables within the equation, outlining the key factors contributing to the signal processing process.

4.2.2. High-pass filter

High-pass filters are frequently employed in EEG and ERP studies to enhance the statistical power of data. They serve the purpose of mitigating variations in voltage, including skin potentials and gradual voltage offsets [25]. This type of electric filter functions by allowing signals possessing frequencies above the cut-off frequency to pass through while simultaneously attenuating signals with frequencies below this threshold. Notably, high-pass filters exhibit reduced susceptibility to the edge artifacts issue, which can lead to inaccuracies in the calculation of filtered values, especially after the EEG recording. As a result, high-pass filters are commonly administered to the EEG recording before the EEG Epoching step, during which specific time windows are extracted from the continuous single-trial EEG data. This strategic application of high-pass filters serves to optimize data quality and reliability in EEG and ERP investigations. A high-pass filter with a cutoff frequency of 0.1 Hz and an attenuation of 12 dB/octave is applied to the raw Brain Signal in this study.

4.2.3. Ensemble averaging

Ensemble averaging stands as a widely adopted technique in ERP analysis, primarily serving the purpose of noise reduction and the amplification of ERP sub-component signals. At its core, this approach involves the precise alignment and subsequent averaging of multiple ERP waveforms that are time-locked to a specific event of interest. To effectively isolate and extract components such as the P300, N170, and N400 from event-related potential (ERP) waveforms through ensemble averaging, a technique known as time-domain averaging is typically employed. This method entails the meticulous alignment of multiple trials associated with the same type of stimulus, followed by the collective averaging. This process is instrumental in elevating the signal-to-noise ratio and facilitating the extraction of the intended ERP component. The general equation for ensemble averaging can be succinctly expressed as follows:

$$\mathrm{ERP}_{\mathrm{avg}}(t) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{ERP}_i(t) \tag{2}$$

Where $\mathrm{ERP}_{\mathrm{avg}}(t)$ is the averaged ERP waveform at time point t. n is the total number of trials. $\mathrm{ERP}_i(t)$ represents the ERP waveform of the i-th trial at time point t. By repeating this process for all time points, a filtered ERP waveform that highlights the component of interest while reducing random variability can be obtained.

4.2.4. Low-pass filter

In contrast to high-pass filters, low-pass filters operate inversely, attenuating brain signals characterized by frequencies lower than the specified cut-off frequency while allowing signals exhibiting frequencies higher than the cut-off frequency to pass through. In the context of ERP studies, low-pass filters find application in the suppression of noise artifacts, including the noise line and EMG interference within the data [25]. It is

pertinent to employ a low-pass filter when working with averaged EEG or ERP waveforms, particularly due to the reduced impact of edge artifacts on shorter waveforms. A low-pass filter with a cut-off frequency of 30Hz and 24db attenuation is applied to the ERP waveform. For optimal filtering performance and the enhancement of spectral resolution, a finite impulse response (FIR) low-pass filter, implemented with a Hamming window as the window design method, is often utilized [10]. The Hamming window, an extension of the Hann window, presents distinctive characteristics, including a heightened side lobe and a more gradual fall-off rate, as compared to the Hann window, as illustrated in Equation (3).

$$w(n) = 0.54 - 0.46 \cos\left(2\pi \frac{n}{N}\right) \tag{3}$$

Where $w(n)$ represents the window coefficient, N corresponds to the total number of signals encompassed within the window frame, and n denotes the input signal. These elements collectively indicate the fundamental parameters and variables within the equation, outlining in the signal processing process.

4.2.5. Baseline correction

ERP signals exhibit a time-resolved nature, which implies the presence of temporal drifts or vertical offsets in the electrical signals. These variations in voltage levels over time often result from a range of factors, both internal and external, such as fluctuations in skin hydration and static changes in electrode conditions during the data collection process. To mitigate the influence of these offsets and drifts on the integrity of ERP signals, a fundamental preprocessing technique known as baseline correction is employed in ERP analysis. The core principle behind baseline correction involves the removal of these temporal drifts and offsets from the recorded brain signals. This is achieved by computing the average voltage level during the pre-stimulus interval and subsequently subtracting this value from every data point within the ERP waveform. The application of baseline correction serves the purpose of minimizing variance in the data while effectively segregating the stimulus-induced brain activity from any preexisting neural activity that may have been present before the onset of the stimulus.

4.3. Correlation and distance-based approach

The Spearman's Rank Correlation Coefficient, in conjunction with Euclidean Distance, has been employed to formulate a framework for defining cognitive performance with a focus on the P300, N170, and N400 ERP sub-components. Spearman's Rank Correlation Coefficient serves to quantify the strength and direction of the monotonic relationships that exist between various variables, while Euclidean Distance is utilized to measure the degree of dissimilarity or similarity between vectors representing these ERP sub-components. This proposed approach offers a comprehensive evaluation of cognitive performance based on the P300, N170, and N400 sub-components, encompassing both the interrelationships and the spatial separation. By incorporating

both the correlations between these sub-components and the spatial distances, this integrated approach yields a more holistic assessment of cognitive performance. The correlation between the amplitudes of the target waveforms can be computed using Equation (4).

$$\rho = 1 - \frac{6 \cdot \sum d^2}{n \cdot (n^2 - 1)} \tag{4}$$

where ρ is the correlation coefficient, d is the difference in ranks of corresponding observations, n is the number of observations (e.g., trials or participants). To provide a more specific representation of the variables in the equation for the sum of squared differences between the ranks of amplitudes, in this case, the variables can be denoted as follows:

- $\mathrm{P300}_{\mathrm{ranks}}$: The ranks of P300 across the EEG channels.
- $\mathrm{N170}_{\mathrm{ranks}}$: The ranks of N170 across the EEG channels.
- $\mathrm{N400}_{\mathrm{ranks}}$: The ranks of N400 across the EEG channels.

To calculate the sum of squared differences between the ranks of these variables, the equation is modified as Equation (5).

$$\sum d^2 = \sum (P300_{\mathrm{ranks}} - N170_{\mathrm{ranks}})^2 + \sum (P300_{\mathrm{ranks}} - N400_{\mathrm{ranks}})^2 + \sum (N170_{\mathrm{ranks}} - N400_{\mathrm{ranks}})^2 \tag{5}$$

The actual computation of the sum of squared differences would involve substituting the specific ranks of the variables into the equation and performing the summation. As the specific measure of learning performance is used, the experimental design will impact the interpretation of the correlation results. The strength of the correlation in the correlation coefficient (ρ) ranges from -1 to 1. A correlation coefficient of 1 indicates a perfect positive correlation, while a coefficient of -1 indicates a perfect negative correlation and a coefficient of 0 indicates no correlation. The significance of the correlation can be determined using a hypothesis test with a p-value threshold (e.g., p ¡ 0.05). Euclidean distance is applied to measure the correlation between ERP waveforms and human learning performance. The basic idea is to calculate the Euclidean distance between two targeted ERP sub-components and use it as a measure of the similarity. A smaller Euclidean distance indicates a stronger correlation between the two waveforms as in Equation (6).

$$d(x, y) = \sqrt{\sum (x_i - y_i)^2} \tag{6}$$

where $d(x, y)$ is the Euclidean Distance between two waveforms x and y. x_{i} and y_{i} are the amplitude values of the i-th sample in waveforms x and y, respectively.

The Euclidean Distance can be computed as the square root of the sum of squared differences between the amplitudes of corresponding components across EEG channels as follows:

- $P300_i$: P300 amplitudes across EEG channels.
- $N170_i$: N170 amplitudes across EEG channels.
- $N400_i$: N400 amplitudes across EEG channels.

The specific computation of Euclidean Distance for the amplitude of the ERP sub-components is presented as Equation (7).

$$d_{ERP} = \sqrt{\sum(P300_i - N170_i)^2 + \sum(P300_i - N400_i)^2 + \sum(N170_i - N400_i)^2} \tag{7}$$

Where d_{ERP} denotes the squared differences between the targeted ERP sub-components. The $P300_{\mathrm{i}}$, $N170_{\mathrm{i}}$, and $N400_{\mathrm{i}}$ are the amplitudes at index i in the respective vectors. Assigning weights to each component of ERP waveforms (P300, N170, N400) based on the importance or relevance is subjective and can vary depending on the specific research question or context. However, this study provides a general approach for assigning weights W_1, W_2, and W_3 to each component (P300, N170, N400) based on the importance or relevance (normalized to a scale of 0 to 1):

- $W_1 = 0.4$
- $W_2 = 0.3$
- $W_3 = 0.3$

These weighted ERP waveforms reflect the assigned weights based on the important criteria by multiplying the Spearman's Rank Correlation Coefficient by the Euclidean Distance, weighted by the respective weights for each component. The weighted correlation value is calculated for each component using the following Equations:

$$P300_{wcc} = \rho \cdot d_{ERP} \cdot W_1 \tag{8}$$

$$N170_{wcc} = \rho \cdot d_{ERP} \cdot W_2 \tag{9}$$

$$N400_{wcc} = \rho \cdot d_{ERP} \cdot W_3 \tag{10}$$

Where WCC denotes the Weighted Correlation Component for each component. The weighted correlation values are summed up across the ERP sub-components to obtain an overall measure of cognitive performance. The proposed methods combined approach considers the correlation between ERP sub-components (P300, N170, N400) using Spearman's Rank Correlation Coefficient and incorporates the spatial distance using Euclidean Distance. A positive correlation suggests that there is a consistent relationship between the ranks of the P300, N170, and N400 components across the brain channels. Higher ranks in one component are associated with higher ranks in the other components, indicating a similar pattern of response across the channels. The magnitude of the positive correlation coefficient would indicate the strength of the relationship. If the correlation coefficient is close to 1, it suggests a strong positive

correlation, meaning that the ranks of the components are closely related and tend to increase or decrease together. If the correlation co-efficient is closer to 0, it indicates a weaker positive correlation. In contrast, the negative correlation coefficient suggests that there is a negative relationship between the ranks of the P300, N170, and N400 components across the Brain channels. This means that ranks in one component tend to be associated with lower ranks in the other components. By applying appropriate weights, this method can capture the relative importance of each component in determining cognitive performance.

5. Experimental results

The ERP sub-components need to be extracted from raw Brain Signals and preprocessed equally before being applied to the Correlation and Distance-Based approach. Initially, a FIR high-pass filter with a cutoff frequency of 0.1 Hz and an attenuation of 12 dB/octave is applied to the continuous single trial Brain signal data. This high-pass filter effectively eliminates noise, optimizing the quality of the Brain Signal data. The effects of the filter on an EEG channel are shown in Figure 3, highlighting the successful noise reduction.

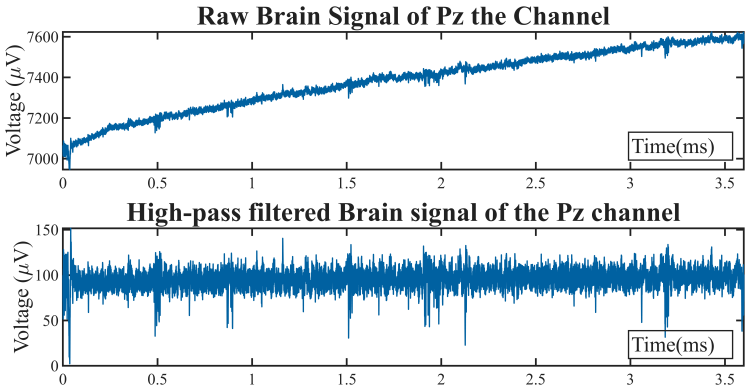


Figure 3. High-Pass Filtered Brain Signal Data of Pz Channel

The high-pass filtered raw Brain Signal is divided into 150 separate trials, each with a 2,400 milliseconds split into 1000 pre-stimulus and 1400 post stimulus duration. These trials are then averaged into a single trial using time-domain averaging to extract the Event-Related Potential signals. This process enhances the signal-to-noise ratio and ensures that the extracted ERP sub-components are representative of cognitive activity.

Table 2 presents the numerical values corresponding to the data points of the Event-Related Potential (ERP) recorded across the Electroencephalogram (EEG) channels within a representative sample following the time domain averaging. Subsequently, a low-pass filter with a cut-off frequency of 30 Hz and 24 db attenuation

is applied to the ERP signal across the Brain channels, reducing the impact of noise and further refining the signal quality.

Table 2
Filtered ERP Data

Time [ms]	Amplitude [μV]									
	Pz	P3	P4	P7	P8	Oz	O1	O2	Cz	CPz
1	95.7	74.4	68.5	–123.7	50.9	42.9	–100.1	77.4	69.8	92.4
2	95.6	74.4	68.4	–123.5	51.0	42.8	–100.2	77.2	69.8	92.4
3	95.6	74.5	68.3	–123.1	51.1	42.6	–100.3	77.0	69.8	92.4
...	...	...	...	...	...	...	...	...	...	...
2399	95.8	74.5	68.5	–123.6	51.3	43.2	–100.1	77.5	69.9	92.3
2400	95.9	74.4	68.6	–123.4	51.4	43.3	–100.0	77.6	69.8	92.3

In addition, a baseline correction will be enacted on the data, wherein a pre-stimulus duration of 1,000 milliseconds will be employed. This correction involves computing the average of the first 1,000 data points in the ERP signal and subtracting this computed value from the entire waveform within each EEG channel. Afterward, 75 evenly distributed data points from the time windows of 250–350 ms, 120–200 ms, and 300–500 ms are extracted to take the amplitudes of the P300, N170, and N400 ERP sub-components respectively, as specified in Table 1. Only 75 data points are selected because it is the greatest number evenly spaced of data points extractable from each ERP sub-component time window. Figure 4 shows the extracted amplitudes of the ERP sub-components from the 10 channels of a sample, indicating the presence of the components of interest.

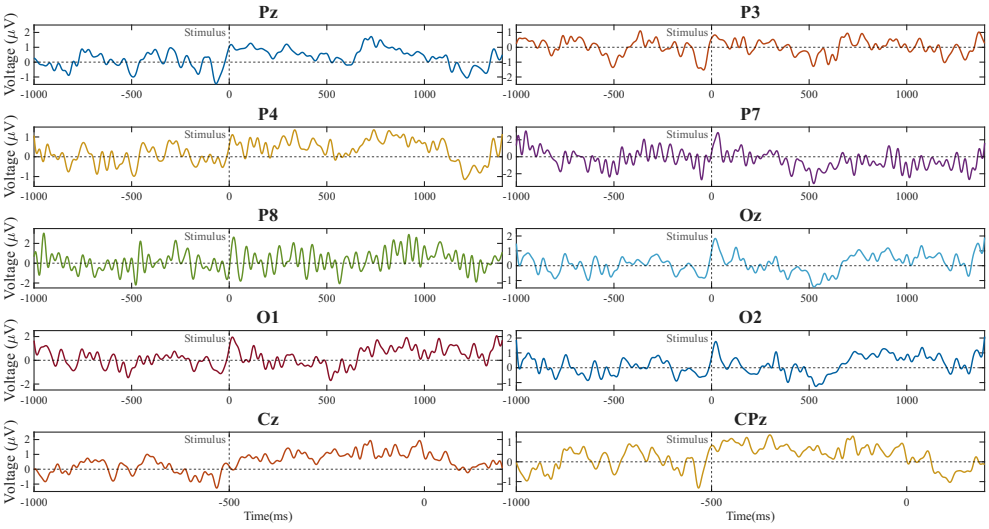


Figure 4. Preprocessed ERP Waveforms Across the Brain Channels

Finally, ensemble averaging is applied to the extracted amplitudes of the ERP sub-components by taking the average of the amplitudes of the 10 channels at each time point. This ensemble averaging provides a consolidated representation of the ERP components and enhances the statistical significance of the analysis. The result is shown in Table 3, indicating the averaged amplitudes of the P300, N170, and N400 ERP sub-components.

Table 3

Ensemble-Averaged Amplitudes of P300, N170 and N400 ERP Sub-Components

P300		N170		N400	
Time [ms]	Activity [μV]	Time [ms]	Activity [μV]	Time [ms]	Activity [μV]
250	0.3930	120	0.7729	300	0.6764
251	0.4056	121	0.7933	303	0.6525
253	0.4242	122	0.8110	305	0.6328
...	...	...	...	...	...
349	0.2277	199	0.2098	497	-0.0548
350	0.1969	120	0.2156	500	-0.0368

To analyze the Event-Related Potential components with the Correlation and Distance-Based approach, the 75 data points of each ERP sub-component are rounded to 2 decimal places, standardizing the data for further analysis. The data points are sorted in ascending order and given a rank based on the values compared to other data points. This ranking process is essential for quantifying the relationships between ERP components. The ranked amplitudes are then used to calculate the sum of squared rank differences between the P300, N170, and N400 ERP sub-components (see Fig. 5). The resulting sum of squared rank differences, along with the number of ranked data points, is employed to compute Spearman's Rank Correlation Coefficient, utilizing Equation (7).

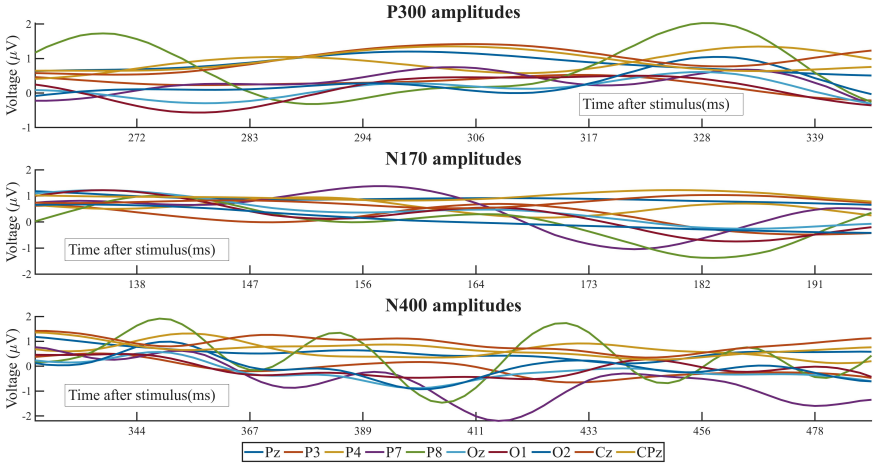


Figure 5. Amplitudes of P300, N170, and N400 ERP Sub-Components

For the distance-based part of the method, the square root of another sum of squared differences between the sorted amplitudes of the ERP sub-components is taken to obtain the Euclidean Distance between the amplitudes of the ERP sub-components. This Euclidean Distance measurement provides insight into the spatial relationships between the ERP components. Finally, the weighted correlation coefficients of each ERP sub-component are calculated by multiplying the Spearman's Rank correlation coefficient with the Euclidean Distance and the assigned weights for the respective ERP sub-component, following Equations (8), (9), and (10) from Section 4.3. This weighted correlation approach allows for a comprehensive assessment of the contributions of each ERP sub-component to cognitive performance. The descriptive statistics of the results obtained from the ERP analysis using the Correlation and Distance-Based approach are shown in Table 4.

Table 4
Results of the Correlation Coefficient and Distance-Based approach

Variable	Mean	Standard deviation	Variance	Confidence interval
Correlation Coefficient (ρ)	0.99	0.01	0.0002	0.01
Euclidean Distance (dERP)	5.50	2.70	7.30	0.93
P300wcc	2.18	1.08	1.16	0.77
N170wcc	1.63	0.80	0.65	0.58
N400wcc	1.63	0.81	0.65	0.58

The examination of correlation coefficients displayed in Table 4 unveils profound insights into the interrelationships among the P300, N170, and N400 ERP sub-components and the impact on cognitive performance. A Spearman's rank correlation coefficient of 0.99, accompanied by a low standard deviation of 0.01, accentuates a positive correlation between the amplitudes of these sub-components and cognitive performance. This implies that greater amplitudes of the P300, N170, and N400 sub-components are indicative of enhanced cognitive performance.

Complementing this insight, the analysis of Euclidean Distance data corroborates the observation. With a mean value of 5.50, a standard deviation of 2.70, and a Confidence Interval of 0.93, the relatively diminutive Euclidean Distance underscores the spatial proximity of data points associated with these ERP sub-components. This suggests that the P300, N170, and N400 sub-components share analogous predictive capacities concerning cognitive performance. The combined evidence from correlation coefficients and Euclidean Distance data underscores a positive relationship among these sub-components, establishing a strong case that higher amplitudes are reflective of superior cognitive performance.

Moreover, the results of paired two-sample t-tests serve to fortify these findings. The p-value of 0.0001 derived from the comparison of $P300_{\text{wcc}}$ with the ERP sub-components underscores the statistical preeminence of $P300_{\text{wcc}}$ in cognitive perfor-

mance assessment, thus underscoring P300 as the most reliable ERP sub-component for this purpose. Conversely, the p-value of 0.999 arising from the paired two-sample t-test contrasting $N170_{\text{wcc}}$ and $N400_{\text{wcc}}$ implies the absence of a substantial distinction in the efficacy in the evaluation of cognitive performance, reaffirming the equal importance of the N170 and N400 ERP sub-components within this context.

In summation, the systematic scrutiny of correlation coefficients, Euclidean Distance data, and paired t-tests collectively reiterates the positive correlation between the P300, N170, and N400 ERP sub-components and cognitive performance. This evidence strongly suggests that elevated amplitudes within these sub-components correlate with superior cognitive performance. These findings offer invaluable guidance in prioritizing the P300 ERP sub-component in cognitive performance research, while simultaneously acknowledging the equal significance of N170 and N400 in the evaluation of cognitive performance.

6. Discussion and conclusion

6.1. Discussion

The identification of cognitive performance in humans using event-related potentials (ERP) has been a topic of extensive research for several decades. In recent years, significant advancements have been made in the field, particularly in the development of advanced techniques and methods that enable more precise and dependable measurements of cognitive performance through ERP analysis. Our experimental results align with these advancements by demonstrating the effectiveness of filtering and preprocessing techniques in enhancing the quality of Brain signal data, which is crucial for accurate ERP analysis.

A notable recent development involves the application of machine learning algorithms to classify ERP waveforms. Deep learning and support vector machines, among other machine learning algorithms, have shown promising outcomes when utilized to analyze ERP data. These algorithms possess the capability to identify patterns and correlations within ERP data that may not be readily discernible to human observers [32] [3]. This is particularly relevant to our study, where the Correlation and Distance-Based approach was used to analyze the ERP sub-components, providing a robust method for assessing cognitive performance.

Moreover, recent progress in ERP research involves the utilization of high-density EEG systems, which enable the recording of ERP data with superior spatial resolution. This advancement offers the potential to extract more detailed information about the neural processes underlying cognitive performance [33], enhancing the significance of our findings related to the P300, N170, and N400 sub-components. The integration of multimodal imaging techniques, such as combining EEG with functional magnetic resonance imaging (fMRI), has facilitated a more comprehensive understanding of the neural mechanisms associated with cognitive performance [5, 9], further supporting the potential applications of our study.

In addition to investigating cognitive performance, recent studies have focused on utilizing ERP data to develop brain-computer interfaces (BCIs) for individuals with disabilities [38]. The utilization of ERP data in the development of BCIs has demonstrated encouraging results, enabling individuals with disabilities to control devices such as prosthetic limbs and communication devices [13]. These advancements highlight the broader implications of ERP research, extending beyond cognitive performance assessment to practical applications that can significantly impact individuals' lives.

6.2. Conclusion

The correlation analysis conducted in this study has yielded valuable insights into the associations between the P300, N170, and N400 event-related potential (ERP) sub-components and cognitive performance. These ERP sub-components, which are linked to various cognitive domains, have emerged as potential biomarkers for evaluating cognitive abilities in educational and training settings. The strong positive correlation, as indicated by Spearman's rank correlation coefficient of 0.99 and a low standard deviation of 0.01, underscores the relationship between the amplitudes of these sub-components and cognitive performance.

The high-pass filtering, and time-domain averaging applied to the raw Brain signal data effectively enhanced the signal-to-noise ratio, allowing for clear extraction of ERP sub-components. The subsequent application of low-pass filtering and baseline correction further refined the signal quality, enabling precise measurement of the P300, N170, and N400 amplitudes. The Euclidean Distance analysis substantiates the strong correlation observed, revealing a relatively small distance between the ERP sub-components, which indicates similar predictive power regarding cognitive performance.

The results of paired two-sample t-tests reinforce these findings. The p-value of 0.0001 in the comparison between $P300_{\text{wcc}}$ and the N-group ERP sub-components highlights the statistical superiority of $P300_{\text{wcc}}$ in assessing cognitive performance. In contrast, the paired two-sample t-test between $N170_{\text{wcc}}$ and $N400_{\text{wcc}}$, with a p-value of 0.999, indicates no significant difference in their effectiveness in assessing cognitive performance. This underscores the equal importance of the N170 and N400 ERP sub-components in this specific context.

In summary, the study's results demonstrate a positive correlation between the P300, N170, and N400 ERP sub-components and cognitive performance, suggesting that higher amplitudes of these sub-components are indicative of superior cognitive performance. The study emphasizes the significance of the P300 sub-component while also recognizing the necessity of N170 and N400 in the assessment of cognitive performance. These findings offer valuable insights for future research endeavors and practical applications in the field of cognitive performance assessment.

6.3. Limitations and future works

There are several limitations to studies that aim to identify cognitive performance using ERP. One of the main limitations is that the quality of the ERP signal is influenced by various factors such as age and gender, mental state, and the type of EEG equipment used for the participant. Additionally, the process of collecting and analyzing ERP data is time-consuming and requires significant expertise, which can limit the practicality and scalability of the approach. Besides, ERP signals may be influenced by other factors besides cognitive performance. For example, external stimuli such as sounds or visual cues can trigger certain ERP sub-components, which may not necessarily reflect cognitive performance. Moreover, different individuals may have varying levels of baseline ERP activity, which can affect the interpretation of ERP signals. Furthermore, there is a lack of consensus on the precise relationship between ERP sub-components and cognitive performance. While some studies have reported strong correlations between specific ERP sub-components and cognitive abilities, others have reported weak or inconsistent relationships. The complex nature of cognitive performance, which involves multiple cognitive processes and neural networks, makes it difficult to identify specific ERP sub-components that can accurately predict cognitive performance. Finally, there are ethical and privacy concerns related to the use of ERP signals for cognitive assessment. The collection and analysis of brain activity data may raise concerns about the confidentiality and privacy of the individual, as well as potential discrimination based on cognitive ability. Despite these limitations, the use of ERP signals for cognitive assessment shows promises and has the potential to provide valuable insights into cognitive performance. Ongoing research is focused on addressing these limitations and developing more accurate and reliable methods for identifying cognitive performance using ERP signals.

Acknowledgements

The authors extend the sincere gratitude to Assoc. Prof. Daniel Baker, a distinguished senior lecturer at the esteemed York Biomedical Research Institute (YBRI) located at the University of York, England, for his valuable insights and the dataset utilized in this study. Moreover, this study is supported by the Payap University.

Ethical approval

This study did not involve any human or animal subjects, as it utilized an open-source event-related potential (ERP) dataset obtained through contrast discrimination tasks at the York Biomedical Research Institute (YBRI), University of York, England. Therefore, ethical approval, consent to participate, and consent to publish were not applicable.

Funding

No external funding was received for this study. The research was conducted without financial support from any external sources.

Availability of data and materials

The dataset utilized in this study is an open-source event-related potential (ERP) dataset obtained through contrast discrimination tasks at the York Biomedical Research Institute (YBRI), University of York, England. The dataset can be accessed through the official channels of the YBRI. For inquiries regarding data access, please contact Associate Professor Daniel Baker, a distinguished senior lecturer at the YBRI, who provided the necessary permissions for the implementation of the dataset.

References

[1] Baker D.H.: Raw ERP Data in CSV Format, osf.io/xu87h, 2019.

[2] Bentin S., Allison T., Puce A., Perez E., McCarthy G.: Electrophysiological Studies of Face Perception in Humans, *Journal of Cognitive Neuroscience*, vol. 8(6), pp. 551–565, 1996. doi: 10.1162/jocn.1996.8.6.551.

[3] Borra D., Bossi F., Rivolta D., Magosso E.: Deep learning applied to EEG source-data reveals both ventral and dorsal visual stream involvement in holistic processing of social stimuli, *Scientific Reports*, vol. 13(1), 7365, 2023. doi: 10.1038/s41598-023-34487-z.

[4] Chou W.C., Duann J.R., She H.C., Huang L.Y., Jung T.P.: Explore the Functional Connectivity between Brain Regions During a Chemistry Working Memory Task, *PloS One*, vol. 10(6), e0129019, 2015. doi: 10.1371/journal.pone.0129019.

[5] Ciccarelli G., Federico G., Mele G., Di Cecca A., Migliaccio M., Ilardi C.R., Alfano V., Salvatore M., Cavaliere C.: Simultaneous real-time EEG-fMRI neurofeedback: A systematic review, *Frontiers in Human Neuroscience*, vol. 17, 2023. doi: 10.3389/fnhum.2023.1123014.

[6] Cremone-Caira A., Vaidyanathan A., Hyatt D., Gilbert R., Clarkson T., Faja S.: Test-retest reliability of the N2 event-related potential in school-aged children with autism spectrum disorder (ASD), *Clinical Neurophysiology*, vol. 131(2), pp. 406–413, 2020. doi: 10.1016/j.clinph.2019.09.024.

[7] Dokmanic I., Parhizkar R., Ranieri J., Vetterli M.: Euclidean Distance Matrices: Essential theory, algorithms, and applications, *IEEE Signal Processing Magazine*, vol. 32(6), pp. 12–30, 2015. doi: 10.1109/MSP.2015.2398954.

[8] Duncan-Johnson C.C., Donchin E.: On Quantifying Surprise: The Variation of Event-Related Potentials with Subjective Probability, *Psychophysiology*, vol. 14(5), pp. 456–467, 1977. doi: 10.1111/j.1469-8986.1977.tb01312.x.

[9] Ebrahimzadeh E., Saharkhiz S., Rajabion L., Oskouei H.B., Seraji M., Fayaz F., Saliminia S., *et al.*: Simultaneous electroencephalography-functional magnetic resonance imaging for assessment of human brain function, *Frontiers in Systems Neuroscience*, vol. 16, 934266, 2022. doi: 10.3389/fnsys.2022.934266.

[10] Emerson: Software Filtering: Windowing – General Analog Concepts. www.ni.com/en/support/documentation/supplemental/06/software-filtering–windowing—general-analog-concepts.html. 2023.

[11] Finisguerra A., Borgatti R., Urgesi C.: Non-invasive Brain Stimulation for the Rehabilitation of Children and Adolescents with Neurodevelopmental Disorders: A Systematic Review, *Frontiers in Psychology*, vol. 10, 135, 2019. doi: 10.1155/2023/5612061.

[12] Folstein J.R., Van Petten C.: Influence of cognitive control and mismatch on the N2 component of the ERP: A review, *Psychophysiology*, vol. 45(1), pp. 152–170, 2008. doi: 10.1111/j.1469-8986.2007.00602.x.

[13] Galiotta V., Quattrociocchi I., D'Ippolito M., Schettini F., Aricò P., Sdoia S., Formisano R., *et al.*: EEG-based Brain-Computer Interfaces for people with Disorders of Consciousness: Features and applications. A systematic review, *Frontiers in Human Neuroscience*, vol. 16, 2022. doi: 10.3389/fnhum.2022.1040816.

[14] Harwood V., Kleinman D., Puggioni G., Baron A.: The P300 event-related potential predicts phonological working memory skills in school-aged children, *Frontiers in Psychology*, vol. 13, 2022. doi: 10.3389/fpsyg.2022.918046.

[15] Hauke J., Kossowski T.: Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data, *Quaestiones Geographicae*, vol. 30(2), pp. 87–93, 2011. doi: 10.2478/v10117-011-0021-1.

[16] Hileman C.M., Henderson H., Mundy P., Newell L., Jaime M.: Developmental and Individual Differences on the P1 and N170 ERP Sub-Components in Children with and without Autism, *Developmental Neuropsychology*, vol. 36(2), pp. 214–236, 2011. doi: 10.1080/87565641.2010.549870.

[17] Horvath A., Szucs A., Csukly G., Sakovics A., Stefanics G., Kamondi A.: EEG and ERP biomarkers of Alzheimer's disease: a critical review, *Frontiers in Bioscience (Landmark Edition)*, vol. 23(2), pp. 183–220, 2018. doi: 10.2741/4587.

[18] Intriligator J., Polich J.: On the relationship between EEG and ERP variability, *International Journal of Psychophysiology*, vol. 20(1), pp. 59–74, 1995. doi: 10.1016/0167-8760(95)00028-Q.

[19] Itier R.J., Taylor M.J.: N170 or N1? Spatiotemporal Differences between Object and Face Processing Using ERPs, *Cerebral Cortex*, vol. 14(2), pp. 132–142, 2004. doi: 10.1093/cercor/bhg111.

[20] Kappenman E.S., Luck S.J. (eds.): *The Oxford Handbook of Event-Related Potential Components*, Oxford University Press, 2011. doi: 10.1093/oxfordhb/9780195374148.001.0001.

[21] Karamacoska D., Butt A., Leung I.H.K., Childs R.L., Metri N.J., Uruthiran V., Tan T., *et al.*: Brain Function Effects of Exercise Interventions for Cognitive Decline: A Systematic Review and Meta-analysis, *Frontiers in Neuroscience*, vol. 17, 1127065, 2023. doi: 10.3389/fnins.2023.1127065.

[22] Kayser J., Tenke C.E.: Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: II. Adequacy of low-density estimates, *Clinical Neurophysiology*, vol. 117(2), pp. 369–380, 2006. doi: 10.1016/j.clinph.2005.08.033.

[23] Kutas M., Federmeier K.D.: Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP), *Annual Review of Psychology*, vol. 62, pp. 621–647, 2011. doi: 10.1146/annurev.psych.093008.131123.

[24] Lopez K.L., Monachino A.D., Vincent K.M., Peck F.C., Gabard-Durnam L.J.: Stability, change, and reliable individual differences in electroencephalography measures: A lifespan perspective on progress and opportunities, *NeuroImage*, vol. 275, 120116, 2023. doi: 10.1016/j.neuroimage.2023.120116.

[25] Luck S.J.: *An Introduction to the Event-Related Potential Technique*, The MIT Press, 2005.

[26] Lumsden J., Edwards E.A., Lawrence N.S., Coyle D., Munafò M.R.: Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy, *JMIR Serious Games*, vol. 4(2), e11, 2016. doi: 10.2196/games.5888.

[27] Madhulatha T.S.: An Overview on Clustering Methods, *IOSR Journal of Engineering*, vol. 2(4), pp. 719–725, 2012. doi: 10.9790/3021-0204719725.

[28] Mc Ardle R., Del Din S., Galna B., Thomas A., Rochester L.: Differentiating dementia disease subtypes with gait analysis: feasibility of wearable sensors?, *Gait Posture*, vol. 76, pp. 372–376, 2020. doi: 10.1016/j.gaitpost.2019.12.028.

[29] Picton T.W.: The P300 Wave of the Human Event-Related Potential, *Journal of Clinical Neurophysiology*, vol. 9(4), pp. 456–479, 1992. doi: 10.1097/00004691-199210000-00002.

[30] Polich J.: Updating P300: An integrative theory of P3a and P3b, *Clinical Neurophysiology*, vol. 118(10), pp. 2128–2148, 2007. doi: 10.1016/j.clinph.2007.04.019.

[31] Pyc M.A., Rawson K.A.: Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory?, *Journal of Memory and Language*, vol. 60(4), pp. 437–447, 2009. doi: 10.1016/j.jml.2009.01.004.

[32] Salehzadeh R., Soylu F., Jalili N.: A comparative study of machine learning methods for classifying ERP scalp distribution, *Biomedical Physics & Engineering Express*, vol. 9(4), 045027, 2023. doi: 10.1088/2057-1976/acdbd0.

[33] Stoyell S.M., Wilmskoetter J., Dobrota M., Chinappen D.M., Bonilha L., Mintz M., Brinkmann B.H., *et al.*: High Density EEG in Current Clinical Practice and Opportunities for the Future, *Journal of Clinical Neurophysiology: Official Publication of the American Electroencephalographic Society*, vol. 38(2), pp. 112–123, 2021. doi: 10.1097/WNP.0000000000000807.

[34] Toffolo K.K., Freedman E.G., Foxe J.J.: Evoking the N400 Event-related Potential (ERP) Component Using a Publicly Available Novel Set of Sentences with Semantically Incongruent or Congruent Eggplants (Endings), *Neuroscience*, vol. 501, pp. 143–158, 2022. doi: 10.1016/j.neuroscience.2022.07.030.

[35] Whelan R., Barbey F.M., Cominetti M.R., Gillan C.M., Rosická A.M.: Developments in scalable strategies for detecting early markers of cognitive decline, *Translational Psychiatry*, vol. 12(1), 473, 2022. doi: 10.1038/s41398-022-02237-w.

[36] Wlotko E.W., Lee C.L., Federmeier K.D.: Language of the Aging Brain: Event-Related Potential Studies of Comprehension in Older Adults, *Language and Linguistics Compass*, vol. 4(8), pp. 623–638, 2010. doi: 10.1111/j.1749-818X.2010.00224.x.

[37] Woodman G.F.: A brief introduction to the use of event-related potentials in studies of perception and attention, *Attention, Perception & Psychophysics*, vol. 72(8), pp. 2031–2046, 2010. doi: 10.3758/APP.72.8.2031.

[38] Yadav H., Maini S.: Electroencephalogram based brain-computer interface: applications, challenges, and opportunities, *Multimedia Tools and Applications*, vol. 82, pp. 47003–47047, 2023. doi: 10.1007/s11042-023-15653-x.

[39] Yener G., Hünerli-Gündüz D., Yıldırım E., Aktürk T., Başar-Eroğlu C., Bonanni L., Del Percio C., *et al.*: Treatment effects on event-related EEG potentials and oscillations in Alzheimer's disease, *International Journal of Psychophysiology*, vol. 177, pp. 179–201, 2022. doi: 10.1016/j.ijpsycho.2022.05.008.

Affiliations

Nyi Nyein Aung
Payap University, Department of Information Technology, International College,
Super-highway Chiang Mai, Lumpang Road, Amphur Muang Chiang Mai, 50000, Thailand,
leonyinyeinaung@gmail.com

Wanus Srimaharaj
Payap University, Department of Information Technology, International College,
Super-highway Chiang Mai, Lumpang Road, Amphur Muang Chiang Mai, 50000, Thailand,
wanus_s@payap.ac.th

Received: 23.01.2024
Revised: 30.05.2024
Accepted: 9.06.2024

Computer Science • 25(4) 2024 https://doi.org/10.7494/csci.2024.25.4.5423

Marcus Hilbrich
Ninon De Mecquenem

MICROSERVICES, A DEFINITION ANALYZED BY ßMACH

Abstract *Managing software artifacts is one of the most essential aspects of computer science. It enables to develop, operate, and maintain software in an engineer-like manner. Therefore, numerous concrete strategies, methods, best practices, and concepts are available. A combination of such methods must be adequate, efficient, applicable, and effective for a concrete project. Eelsewise, the developers, managers, and testers should understand it to avoid chaos. Therefore, we exemplify the ßMACH method that provides software guidance. The method can point out missing management aspects (e.g., the V-model is not usable for software operation), identify problems of knowledge transfer (e.g., how is responsible for requirements), provide an understandable management description (e.g., the developers describe what they do), and some more. The method provides a unified, knowledge-based description strategy applicable to all software management strategies. It provides a method to create a minimal but complete description. In this paper, we apply ßMACH to the microservice concept to explain both and to test the applicability and the advantages of ßMACH.*

Keywords ßMACH, microservices, software artefact management, process model, architectural style

Citation Computer Science 25(4) 2024: 547–578

1. Introduction

Managing software artifacts is one of the most essential aspects of computer science. The question is how to develop, operate, and maintain software [15, 66, 75, 76, 80, 81]. Scientists and industry give different (partial) answers to this question: software process models [1, 24, 45, 69, 70, 73, 74], programming paradigms [4, 7], change management [3, 8, 10, 13, 22, 30, 32, 38, 52, 86], best practices [9, 39–42, 53, 85], and many more [14, 17, 23, 27, 47, 60, 79, 84]. So, to tackle one challenge, we have various solution strategies, methods, description languages, and suggestions present. We neither have a uniform solution strategy nor a description language. For managing concrete software, e.g., in a project [67, p. 1] [70, p. 3,5,14], software developers, managers, testers, and reviewers are deciding and understanding/learning a management strategy. Thus, the open question in concrete is.

How to review if a software artifact management strategy is suitable ahead of the beginning of the management?

In principle, a management strategy has to cover at least all relevant aspects of software management. So, the V-Model is insufficient if a project requires maintaining and operating software. A software management strategy should not cover additionals, to be minimal. So, a change management process is insufficient if the decision to create software is already done. Knowledge transfer is an additional challenge. If a Kanban-based project decides on an explicit requirement engineering, the results of the requirement engineering are included in the development process. Therefore, the requirement engineer can deliver a programmer understandable description or participate in the development process. Requirements are not present in a Scrum process. Such a process requires user stories. They also describe the same functional properties of software artifacts.

The examples describe the overall challenges. The different management strategies cover varied aspects, but a project requires managing exactly the project's aspects. All involved persons have to understand their role in the overall process. Creating and transferring (by persons, documents, and software artifacts) has to be explicit to enable management. The different software management strategies use different terms and languages, which makes them hard to understand and combine.

Our solution is a uniform, complete, minimal, and easy-to-apply software guidance. It applies to all management strategies and their combinations. ßMACH handled the challenge that an a priori a unification of the languages used in computer science is not possible [44, 91] [72, p. 319]. Nevertheless, it is possible to describe the knowledge. Examples are the description of ontology's of engineering [75, 76], the work by Popper [63, 64], or other general categorizations [76].

The ßMACH method [33, 37] provides a unification of the aspects based on a unification of management strategies, an extraction of relevant management aspects, and a systematization of these aspects. As a result, the ßMACH method enables a software developer to check and specify a concrete management strategy with minimal effort. The management strategy becomes easily understandable to programmers,

project managers, engineers, testers, etc. The ßMACH method enforces a description of all relevant aspects of software management and identifies missing ones enforce completeness. It enables review strategies, fosters improvements, is a starting point for academic discussions, and is the basis for systematic comparisons of management strategies. Therefore, it can describe all kinds of management strategies in a uniform knowledge-based language and avoids chaotic processes.

In this paper, we exemplify the ßMACH method to explain it, provide its advantages, and test its applicability. Therefore, we have to decide on a management strategy. Scrum, V-model, or Kanban are describing such management strategies. They are easily describable by the ßMACH method, so we decide on a more challenging task.

We decided on microservices. It is a concept, not a management strategy. Microservices are an answer to various scalability challenges. They enable large and complex systems by scaling the number of services and development teams. Microservices allow self-management and agile processes such as Scrum. Therefore, the developers apply the microservices concept to all parts of the overall system. It requires all developers, managers, software engineers, etc., to understand the microservice concept and to follow it, so the ßMACH-based description in this paper is helpful.

This paper focuses on using the ßMACH method and not creating the ßMACH concept. Therefore, we explain the usage of ßMACH on a toy example and analyze microservices. We decided on microservices because they are well-established and widely used in large-scale industrial applications [55,56,78]. Such systems have proven to be scalable to support several million users. Academics described them at various conferences [16, 28, 29, 46, 51, 54, 77] and discussed them heavily. Thus, microservices are an answer to actual scalability challenges and an object of academic research.

The scalability of microservices is not limited to the user load. Also, the development is scalable. Microservice systems consist of individual microservices [11, 19, 58, 90, 93]. Every service is developed and managed by one team (but a team can have multiple services). The idea of microservices is to keep a team small, as described by Levis and Fowler [49]. Most importantly, the teams and the microservices stay small, even if the overall system can scale. It scales by adding additional microservices and teams.

The small teams provide a set of advantages like less management overhead. The team members are more productive in a flat hierarchy, and agile software management processes are easy to learn. The microservices and teams are independent of each other. Thus, the overall management overhead is reduced. Nevertheless, too many requests based on service communication or other aspects can hinder the productivity of a small team [53]. It is the reason for demanding service and team isolation. It is an essential factor of microservices. Based on the isolation, we call our microservice definition a strong one, as opposed to code size, the number of team members, or the used technology as the basis [2, 5, 57, 82, 89].

In a microservice system, even if the teams are small and self-managed [21,43,70], a minimum set of rules has to be set. The teams should not break the microservice

system, e.g., by building interfaces to other microservices. In addition, all management aspects need a description. Software management includes creating, improving, deploying, and operating software artifacts [74]. Also essential are the documentation and communication of the teams. All members need to understand and agree to the process to avoid conflicts.

ßMACH [33,37] documents and defines a software management process. It aims to check the management strategy to support all relevant management aspects and avoid unnecessary management. ßMACH is minimalistic, based on scientific groundwork and an ontology of key management aspects. It provides engineer-like systematics. In this paper, we give a ßMACH protocol to accomplish all significant aspects of software management in a microservice team. It describes how to create such a protocol. So you can adapt it to your concrete process. As a result, based on the ßMACH protocol, we can demonstrate how the independence of microservices is a solution to many management aspects.

We organize the paper as follows: An introduction to ßMACH in Section 2, to give the fundamentals of the scientific method. In Section 3, the microservice system is defined and explained. Section 4 describes a use case. Then (Section 5), we explain how to fill the ßMACH protocol (perform the ßMACH method). With the ßMACH protocol, we can provide observations on the microservice-based process to manage, and we will analyze microservices (Section 6). We close with a conclusion (Section 7).

2. A short explanation of ßMACH

The ßMACH[1] method is an approach to define and plan a software-management process, e.g., a software development project. Therefore, ßMACH defines the management process, gives additional context (meta-information), and describes how to cover essential aspects of software management (As groundwork, see, e.g., [4,6,8,12,18,20,25,26,47,50,59,62,73,74,87]). The key aspects are based on an ontology of software engineering and software management strategies and are described based on the vocabulary of knowledge management.

The ßMACH protocol consists of three parts: the definition of management processes, meta-information, and descriptions of the key aspects. A team should fill in a protocol for each separate management process. The meta-information defines the team, the filler of the protocol, and additional parts. The definition of *Our Team* and *Cooperating Teams* are essential for understanding the protocol. Our team is the group of persons who directly manage the software. In our example, the team manages one microservice of the overall microservice system. Cooperating teams are other teams that are intensively involved. An example would be a dedicated testing team. In our case study, there is no such team.

[1] Systematic Software Management Approaches Characterization Helper; ß is the German Eszett. You can read and pronounce it as “ss”.

By defining the management process (Fig. A1 and 1), we provide the guidelines to plan the management. It can be short and link to additional documents. For instance, we can reference the Scrum Guide [70] in the case of Scrum-like management. The definition should be easy to understand by the target audience. Usually, this audience is the team. In this case, the readers of the paper. The definition should be in numbered bullet points. So it can efficiently describe the key aspects.

ßMACH defines three groups: *our team*, *cooperating teams*, and *externals*. *Our team* is the group that manages the software, so they have to deliver and operate a microservice (of the microservice system). Also, *our team* defines and learns the management process in the ßMACH protocol. *Cooperating teams* are other teams that *our team* can or must cooperate with. *Our team* can not define how *cooperating teams* work (they do their own management). *Our team* can change the agreements with them during the management process. Such teams are, e.g., teams in the same organization. Because *our team* is working with such teams, ßMACH calls this *Internal*. The last group is called *External*. *Our team* can not directly influence such parties, e.g., contract partners. *Our team* has a defined contract and has to follow it. Another example is a provider of a library or end-users who use the microservice.

The ßMACH protocol organizes the description of the key aspects of software management in a table. Fig. 3 provides an example. The columns define the different aspects of knowledge and information management. It includes who is doing (column *Roles*), what needs to be known to perform the process (column *Process Knowledge*), and how is the product or aim of the process (column *Product Knowledge*). This part of ßMACH follows the idea that a product or artifact is created and managed by actors/persons/roles in a process. [74]. In addition, ßMACH points out if a piece of knowledge is not present at the beginning (column *Demanded Knowledge*). Clarifying which knowledge is required is essential as the process needs to find a solution to acquire it during the management. The last column is called *Process Information*, which defines which information has to be provided by the management process, e.g., working hours for billing, the results of meetings, and delivery protocols.

The rows define the product aspects and the party that influences the aspects. Previously, we gave the type of parties. Our team has external parties (marked by the term *Outside*), and our team has cooperating teams (marked by the term *Inside*). Our team is present by the table and needs to conduct the management process based on the key aspects, given in the table.

The rows in the table represent the *Product Properties*. The artifacts our team has to develop/manage. *Interfaces* are the definition of (technical) interfaces of our artifacts to communicate with other systems. *Dependencies* describe everything our team demands to get from others. *Responsibilities* give what our team needs to provide to others (e.g., based on contracts or regulations). Each of the four aspects is present as internal and external. So, ßMACH defines interfaces to cooperating teams that can be discussed and adapted based on the project needs and fixed interfaces to external parties. These aspects cannot be influenced directly by our team.

The last row is *External Artifacts*. It describes that an artifact is taken from another party and included in our project. It is copied (e.g., use an open-source library). As a result of copying, it is irrelevant whether it is from a cooperating team or external. Nevertheless, an external artifact needs management. Our team needs to know how the artifact works, how we will use it, and other consequences (e.g., based on licensees) the team has to consider.

In addition, ßMACH defines relations of different aspects, abstracted as cells in the table. One aspect can require another, so knowledge transfer or transformation is required. A provides relation expresses that an aspect does not need active management. The aspect is handled/provided as a consequence of managing another one. For example, forbidding the usage of external artifacts provides a solution to all related management aspects by avoiding any need for management.

To exemplify the ßMACH method we provide a toy example in the appendix. The toy example provides additional explanations for all its parts. Thereby, it is possible to look up what to fill in the protocol and have a very simple example of a filled protocol. For this paper, we split the protocol into parts to support printing. The original protocol consists of an A4 page for the definition of the management plus meta-information and A3 pages for the description of the key aspects. We give the management definition in Figure A1, and the meta-information in Figure A2. The table with the key aspects has an initial explanation. We present it in Figure A3. The original protocol provides one table with all key aspects. In this paper, we split it into three parts, presented in Figure A4, A5, and A6. For the toy example, we use blue text to visualize everything we (or the team) filled in. The hints are presented in gray text. You can download the ßMACH protocol at https://doi.org/10.5281/zenodo.10992007 to print it in large or zoom in to read all the details.

The toy example is the following. A company has a magical box to create software. So, your challenge is to find out how to get the box. One team of the company is our team. This team wants to use the software itself. Thus, communication, dependencies, management goals, etc., are extremely reduced compared to a realistic project. The magical box can be interpreted as a simplification of outsourcing. So, payment circumstances, problems with the outsourcing partner, etc., are removed from the example.

The toy examples use the different relations of key aspects. The fields in the column process information (Fig. A5 and A6) are all very similar. Based on the management definition, no information is recorded. In each case, the part 6 of the definition is referenced. Therefore, the similarity is visualized by the same background color.

The toy example provides different examples of require relations. One example is from the field inside dependencies / product knowledge to inside product properties / roles. The dependency describes the need to get the magical box. This box provides the product properties, and someone (a role in the team) needs to get the box from another team of the company. As a result, the dependency requires a role to support resolving the dependency.

An example of provide relations is, e.g., present in the row inside dependencies. The inside dependency is defined by getting the magical box. That is what you need to know to handle the dependency. So, it is in the column product knowledge. This field is resolved by a demand relation described above. In addition, the magical box does not need extended management. Because it is so easy, it does not need an additional/extra role to manage it or a process. Thus, product knowledge provides a solution for field roles and field process knowledge. In this case, it provides a solution because it can be denied to have a role or a process to manage the inside dependencies.

Later on, we describe the filling in of an ßMACH protocol in detail. So, we stay with the toy example as a self-explaining, very simplified example.

3. Strict definition of microservices

The paper aims to analyze microservices with the ßMACH method. Therefore, we defined and described microservices in general, and based on the ßMACH method. We are not describing a concrete microservice project, and we do not deny that those real-world systems need to find compromises between the strict isolation our definition demands and practical circumstances. Thus, we do not describe a concrete, but rather a preparation of a management process. It checks if the microservices concept describes all management aspects defined by ßMACH. For a real-world management process, we would need a more concrete context and an adaption to give missing descriptions in the microservice concept.

Before using ßMACH, we start with the general description: The term microservice is not well-defined: The term is used for SOAs [93] build of small services [61,83], for a realization of an organizational structure [49], as a DevOps concept [49, 90], or as architectural style [11,19,49,58,93]. Our definition focuses on the strict isolation of individual services because isolation can be helpful for management processes [12,71].

In the following, we provide a clear definition of microservices. We used definitions stated before (see also [34–36]), a combination of common definitions and strategies, e.g., [11, 19, 49, 58, 93]. We present our definition as a pattern.

Name Microservices (also called Slice Service Style)
Problems to solve Solves the need for scalability concerning the system load and the number of persons/teams developing the system.
Definition The slice service style is an architectural style where the essential aspects of the system are encapsulated in services (slices, microservices, or vertical services). These services deliver functionality to end-users and have no (or minimal) dependencies on other slices of the system. It includes code-sharing, usage of interfaces, sharing of manpower, and management of, e.g., creation, deployment, and operation.
Consequences Because of the separation of slices to allow scalability, the software process model needs to be adapted or tailored. The definition of slices

influences the overall system and has to be done globally (e.g., up to the design phase of the waterfall model), while the creation and operation of the slices are independent. Thus, the (global) software process model has to support independent software development (e.g., by realizing each slice as a DevOps project) and a design or architectural phase at the beginning.
Drawbacks Because the independence of slices includes teams and persons, the structure of the organization developing the system needs to be aligned. In addition, independence reduces the knowledge transfer of the persons of different slices and hinders common reuse techniques. Especially cross-cutting concerns cannot be managed.

4. Use case

In the following, we present concrete use cases from the development team's perspective. We give examples of how the ßMACH protocol can be helpful in concrete and how it is used by the team. Therefore, we use the microservice example as a basis, but we will also point out differences to a concrete ßMACH protocol.

4.1. The external artifact question

When our development fills in the ßMACH protocol, they get to the row about external artifacts. Microservices, as a concept, do not provide a clear and commonly accepted solution strategy. As a result, our team is pointed to this challenge and needs to make a clear and informed decision. Typical answers are the following:

- To reduce the dependencies on external code, we forbid the usage of external liberties. In the ßMACH protocol, we fill in that no knowledge regarding external artifacts exists and no management process is required. We make it clear to the team members by adding "It is forbidden to use external artifacts." as an item to the management definition.
- To forbid external libraries creates new challenges. Encryption, single sign-on, and logging are forced to be re-implemented. This is a high, additional effort and very error-prone. Using established, well-tested, and continually supported libraries is a solution strategy. In such cases, it demands to know the libraries and check that the licenses are exportable. Integrating the liberties requires checking for security issues and update own microservice on demand. Therefore, it requires adding a process and role to the ßMACH protocol.

The ßMACH concept forces the team to decide how to handle external artifacts. The team decides on a strategy and avoids unwanted problems like unmanaged security issues based on outdated libraries.

4.2. Why not use another microservice?

Let us assume we have a running microservice, and our team operates and maintains it. In this situation, our team gets a new member who proposes to use the other microservices to reduce the code base and increase the functionality.

The team can refer to part 1 of the definition (Fig. 1). Thus, the new member can understand the current situation.

The ßMACH is not written in stone. If the situation changes, the protocol can be updated. In this example, it is discussed to remove part 1 of the definition. As a result, all key aspects referring to this part (Fig. 3) are part of the discussion. In concrete, cooperation with other teams has to be established and managed. It is, e.g., needed to have a plan if another service changes the interface or is temporarily unavailable.

Whether it is more effort or risk to manage the relation to other teams/microservices or to not use their services is a decision of our team. ßMACH demands to describe the plan to foster an informed decision.

4.3. The functionality of the microservice

One of the open questions in our example is the functionality of a singular microservice. In ßMACH, this is mainly a question of product knowledge. To program and maintain the microservice, our team needs the related product properties (see Fig 3). In short, the interfaces the microservice provides to the end-user define the product knowledge (the code base of the microservice). The code needs to implement the realization of the interfaces. The responsibilities (the definition of what our microservice has to provide to the end-user) define the interfaces. A chain of demands relations in the ßMACH protocol (see Fig 3) represents the knowledge transfer. The open question in the protocol is who (which role) provides the responsibilities of our service as a system's concern of the overall microservice system.

In our example, we can not answer the system's concerns at all. The concerns require a concrete system and project. Without knowing the aims, purpose, or business model of the microservice system, we can not answer. A real microservice system example has such information available, at least for the overall microservice system.

The ßMACH protocol we provide in this paper is for the development team of one microservice. Thus, from the viewpoint of this team and the corresponding ßMACH, the knowledge of the partial system's concerns (outside responsibilities) needs to be provided somehow. If we create a ßMACH protocol for another microservice and another team, we encounter the same problem. As a result, we demand additional teams that define the business capabilities of the overall microservice system. In addition, such teams separate the overall system's concerns into individual microservices [31]. To describe such a team in ßMACH is another story. It requires having an overall business strategy [12] and dividing [92] the overall business concerns into individual services.

5. Filling of the ßMACH protocol

We use the microservice definition to describe the filling of the ßMACH protocol. The first step is the discussion of the context information. Then, the definition and the aspects of management are discussed in parallel. The results are Fig. 1, 2, and 3.

5.1. ßMACH context

To fill the ßMACH protocol starts with writing down the context information. This part of the ßMACH protocol defines other parts. So, it is a good starting point. Mostly, the context is very clear and easy to fill. We know the name and the date a priori. It is the first version, so we label it as 1.0.

The ßMACH protocol is filled for a team that manages a microservice, not for the organization that manages the overall microservice system. We call the team Microservice Team A, A to indicate that other teams of this kind exist. The team needs more details than a name, so we added an explanation in Fig. 2. This also describes the artifacts. In the pattern definition (Section 3), the part "This includes code-sharing, usage of interfaces, sharing of manpower, and management of, e.g., creation, deployment, and operation." describes the separation and the artifacts. The part "by realizing each slice as a DevOps project" describes the different teams.

The cooperating teams are mostly independent. Thus, we could define them as external. Also, the teams belong to the same organization. The organization manages the overall microservice project. It argues against a complete disjoin. We use the system border of the microservice system as the external border. The mapping of individual microservices to teams is sufficient to describe it: all teams work on the same microservice system as cooperating teams, even if they are independent. As a result, the context information of the ßMACH protocol is present in the ßMACH protocol in Fig. 2.

5.2. ßMACH Definition and software management aspects

ßMACH is a method to define and discuss a software management process. The ßMACH protocol defines the process required by a software development team. The team is responsible for a microservice. We start with the work packages. Afterward, we describe the management process of our team.

5.2.1. Work package responsibilities

Work packages of ßMACH describe if *our team* is responsible for the development, maintenance, and improvement. The pattern-like definition of microservices (Section 3) mentions development, maintenance, and improvement. Development is called "creation". The "deployment" is a part of the development and/or maintenance (depending on static deployment or the system changes its deployment). The "operation" is at least part of maintenance and can include improvement. The mention of DevOps confirms that all work packages are included in the management process.

Thus, the team is responsible for all work packages, and we check them in the ßMACH protocol (Fig. 1).

5.2.2. Definition of the management process

We have a microservice definition (Section 3), but it is not a ßMACH protocol. We need a description where different parts of the definition are easy to reference. Also, each part should describe one aspect and no mixtures.

To get the definition for ßMACH, the definition from Section 3 is decomposed and recomposed. We can split the first sentence of the pattern-like description into parts that are candidates for the ßMACH definition:

- The naming microservice and the classification as architectural style.
- The representation of system concerns as encapsulated services.
- The services deliver functionality to the end-users.
- Services have no (or minimal) dependencies on each other.

The second sentence describes what is included in the services and is independent of other services:

- Services have a code base.
- Services have interfaces.
- Services have a team (“manpower”).
- Services persist over development, maintenance, and improvement (“creation, deployment, and operation”).

The naming and classification as architectural style do not give the descriptions as needed by the ßMACH protocol. In addition, we can reorder the items in the description of the system and the microservices:

1) The microservice system consists of microservices.
2) Microservices have no (or minimal) dependencies on each other.
3) Microservices represent encapsulated system concerns.
4) Microservices are persisting over development, maintenance, and improvement.
5) Microservices deliver functionality to the end-users.
6) Microservices have interfaces.
7) Microservices have a code base.
8) Microservices have a team.

For the ßMACH protocol, it is only allowed to add needed parts to the management definitions. Thus, it is a reasonable strategy to develop the definition of the management process by answering the questions about the management aspects in the table of the ßMACH protocol. This table provides two separate parts. They are the work the team is not responsible for and the part the team needs to manage directly. In the following, we give both parts.

5.2.3. Responsibilities of the team

The team is responsible for product development, maintenance, and improvement. Thus, we have to cross the fields in the table. So, we finished the rows of product development, product maintenance, and product improvement. We do not have to prepare for other teams to overtake the work. It is typical for DevOps-like strategies.

In the following, we have to provide the descriptions of the software management aspects and fill the table. By filling the table, we have to refer to the parts of the software management concept. As the current starting point, this part of the ßMACH protocol is empty.

We start with interfaces. (There are two rows for interfaces in the table.) We already mentioned interfaces in item (6). The interfaces are inside interfaces in case cooperating teams use the interfaces. It would be a kind of dependence that (2) mostly denies. Thus, the interfaces are mostly used by externals. Externals are the end-users according to (5). Item (3) describes the purpose of the microservice. Because the pattern-like definition does not mention other communication, it needs to be offered by the interfaces. So, inside interfaces can be mostly denied. The product properties of the outside interfaces are a subset of the system concerns. For the ßMACH definition, we combined items (1) and (2) from above to part 1 of the ßMACH definition. In addition, we combine items (3), (5), and (6) from above as part 2 of the ßMACH definition (Fig. 1).

We state that internal interfaces are not present. In short, we deny them. To deny internal interfaces means we can deny the need for product knowledge. The team does not need to know anything about nonexisting interfaces. We denied the other cells in the row, too. There is no need to explore additional knowledge (demanded knowledge), no management process is needed (process knowledge), no one needs to do something (roles), and the team cannot record information about the nonexisting process (process information). In other words, based on the fact that no interfaces exist (product knowledge), no role is needed. In ßMACH, this is a provided relation. Based on the cell product knowledge, other cells in the row are filled/inferred. Provide relations are marked by an arrow and according to the coloring of the right side of the cell, as provided by the filled ßMACH protocol in Figure 3.

To deny product knowledge based on the independence of services can be directly transferred to the rows inside product properties, inside dependencies, and inside responsibilities. The arrow for the provided relations and the coloring of the right part of the cells are the same. We applied the same argumentation to the cell product knowledge in different rows. ßMACH demands to use the same color, in this case (it is not a relation, so the left part of the cell is colored). Based on the same argumentation, we used the same color (Fig. 3). So, we finished the rows inside interfaces, inside product properties, inside dependencies, and inside responsibilities.

Part 2 (Fig. 1) of the ßMACH definition does not only define the product knowledge of outside interfaces. The interface and the concerns define the outside product properties. Part 2 defines the product knowledge of the outside responsibilities by

a subset of the system concerns, too. Thus, all three fields get the same color in the ßMACH protocol (left side of the cell). In addition, we visualize that the three fields depend on each other. The concern of the system presented by the responsibility is best. The interface is just the technical and organizational presentation of the responsibility. So, it is dependent on fulfilling the responsibility. The product property is the realization of the interface. So, the fields of product knowledge in the rows outside product properties, outside interfaces, and outside responsibilities are defined.

The definition of an architectural style (Section 3) does not describe how and by whom the artifacts should be managed. The consequences part of the pattern-like definition is helpful. It gives the tailoring of the software process model. We start with the part that describes that DevOps projects are present for each microservice. It helps to describe additional fields. The DevOps team has to provide all needed roles, and it is small enough to manage itself. It gives the roles and the process knowledge for the rows outside product properties and outside interfaces. Because we gave all descriptions based on the DevOps team, we use the same color for all fields (left part of the cells). We present the result in Figure 3. Now, we add the DevOps team to the ßMACH description as part 3.

The DevOps team knows how to develop and maintain the product. Thus, the DevOps team members have product knowledge (row outside product properties). In other words, the DevOps team builds and maintains the software based on their knowledge/experience (and based on the definition of the interfaces). Microservices do not give an additional knowledge object. Therefore, we extended the product knowledge cell in this row and the demand relation (Fig. 3).

Demanded knowledge (row outside product properties) does not exist. The DevOps team manages the artifacts. The outside interfaces present a definition of the product. It adds two provides relations, so everything is present.

The "consequence" section of the pattern-like definition (Section 3) gives the root of the concerns managed by the team. The product knowledge in the rows outside responsibilities demands it. The separation of system concerns is not described (probably given by another team). Thus, it is demanded knowledge for our team. How to obtain this knowledge is unclear. We cannot name the needed process, roles, and process information. We use question marks and red coloring. Also, the ßMACH definition is extended by part 4. The outside interfaces are (mainly) defined by the outside responsibilities. In the case of a concrete end-user, we would need additional aspects, concretization, and adaptions. The DevOps team handles these interfaces. These are two provides relations.

The column process information is not directly covered by the pattern-like definition (Sec. 3), but the concerns of the system can demand such information (e.g., accounting of used resources to benchmark efficiency). Thus, our team transfers the (description of) product knowledge to process information for the rows of outside responsibilities, outside interfaces, and outside process properties.

The pattern-like definition does not give or define the row outside dependencies (e.g., to use an external service or external artifacts). Both can be demanded or forbidden by the system's concerns. So, a demand relation exists. Otherwise, we can expect that the DevOps team manages artifacts and dependencies (similar to outside product properties and outside interfaces we give them in the already defined color). The product information depends on the system concerns (e.g., for the outside interfaces, the system concerns define also the process information directly, but we do not describe it this way). Our team does not demand additional knowledge. We expect the DevOps team to have the needed skills and knowledge. So, we finished the table of ßMACH descriptions (Fig. 3).

5.3. Filled ßMACH protocol

We separated the parts of the ßMACH protocol. Figure 1 presents the definition, Figure 2 the meta-data, and Figure 3 explains the management aspects.

The ßMACH protocol has only two open, not complete answered cells (Fig. 3). The cells describe the separation of concerns of the overall system to isolated microservices. It is a challenge of microservices [48, 65, 68, 88].

The microservice definition describes many cells, especially product knowledge. The system concerns are a basis, with many relations in the ßMACH protocol. The independence of services enables answering inside related rows.

The architectural style does not fully describe roles and processes. The usage of DevOps answers such questions.

Based on the strict description of microservices (the pattern-like definition, Section 3), we can fill a ßMACH protocol (Section 5.3). Thus, the given definition covers nearly all relevant aspects of software management. The red-colored cells in the ßMACH protocol point out the open challenge of microservices to define independent concerns of the overall system.

The microservice definition (Section 3) holds information not present in the ßMACH protocol. Thus, parts of the definition do not describe software management aspects.

The name of the pattern-like definition is (somehow) represented in the context part of the ßMACH protocol (Fig. 2). The "problem to solve" part is not needed to fill the ßMACH protocol. This part can help to decide whether to use microservices or not. It is not in the scope of the ßMACH method. ßMACH helps to understand if all aspects of managing software artifacts are covered. It is no direct helper to decide to use a specific management method, but it can check different strategies. So, the aims of the pattern-like definition and ßMACH are different.

The "definition" sections of the pattern start with the description in an architectural style. The ßMACH protocol does not cover it. So, it describes a pattern property, like the description itself. The rest of the definition sections cover parts 1 and 2 of the ßMACH definition (Fig. 1).

The "consequences" part of the pattern-like definition gives two pieces of information. First, detail of the service separation (part 1 of the ßMACH definition). Second, the separation of concerns and services has to be realized somehow (part 4 in ßMACH). Third, individual services are realized by DevOps (part 3 in ßMACH).

The drawback section of the pattern-like definition is not represented by ßMACH. It describes problems outside our team, and ßMACH does not represent them. It is an additional different aim of the pattern-like definition and ßMACH.

Work Package Responsibilities:

☑ Finelizing Product Development

☑ Finelizing Product Maintains

☑ Finelizing Product Improvement

Definition of the Management Concept:

1. The microservice system consists of microservices, microservices have no (or minimal) dependences to each other.
2. Microservices represent encapsulated system concerns that are delivered via interfaces to endusers.
3. A microservice is managed by a DevOps team that provides all needed knowledge and manages itself.
4. The separation of system concerns to microservices has to be realized, how to do so is not covered by the microservice concept.

Figure 1. Definition of microservices in the ßMACH protocol. The definition gives the numbers of the bullet points/parts. (See Fig. 3 and https://doi.org/10.5281/zenodo.10992169 for more details)

Context (User/Team/Context Information):

Name of the Filler:	Marcus Hilbrich
Represented Team (Internal Border):	Microservice Team A: one of the teams managing microservices of the overall microservice system
Cooperating Teams (External Border):	All teams working on the same microservice system.
Managed Artifacts:	A subset of the microservices of the overall microservice system.
Date:	2020/06/11
Version of Document:	1.0

Comment:
This ßMACH protocol is based on a microservice definition, to test the ßMACH method and the microservice definition.

Figure 2. Context or meta-data of the definition of microservices in the ßMACH protocol. (See Fig. 3 and https://doi.org/10.5281/zenodo.10992169 for more details)

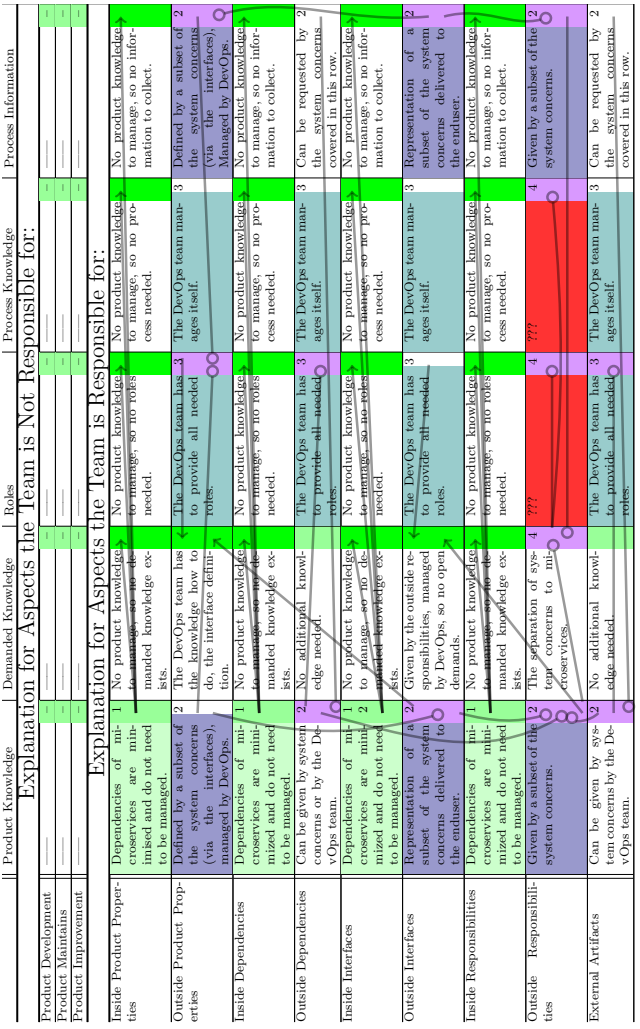


Figure 3. Description of microservices based on the ßMACH protocol. ßMACH describes a set of key aspects. Each cell of the table represents an aspect. The right part of each cell holds the references to the definition in the ßMACH protocol (Fig. 1). ßMACH defines coloring. Based on the management process, we use light green in the right part of a cell for aspects that do not need active management. Active management means that an aspect is realized without a need for action. The darker green indicates that an aspect is also performed without needing active management but is provided by another. We use violet for aspects used or required by additional ones. Such an aspect indicates a special interest. Arrows with a peak-end describe a provides-relation. The aspect on the peak is provided by the other. A round end arrow gives a demand relation. The other aspect needs the one at the rounded end. The left part of the call can be colored, too. If the left part of multiple cells uses the same color, the cell's descriptions are equal or very similar. The described aspects in this figure are all based on the other parts of the ßMACH protocol, provided in Fig. 1 and 2. Download the ßMACH protocol as PDF to zoom in and explore the protocol (https://doi.org/10.5281/zenodo.10992169)

6. Results: learning from the ßMACH Protocol

We investigate the ßMACH protocol (Section 5.3):

1) We start to look for aspects that do not need active management. In the ßMACH protocol, such aspects are marked by a green color on the right part of the cell (Fig. 3). For the strict definition of microservices, the rows for inside aspects do not need active management. The reason is also present. Based on the independence from other services of the same system, no technical (product-based) cooperation with teams of the same microservice system is present. As a result, the other fields in the rows do not need active management because there is nothing to manage. There is no need to manage internal relations, a significant advantage. Fewer communication partners reduce the complexity and the needed team management skills. The team can concentrate on itself and is probably more productive. **ßMACH represents the idea of the strict microservice definition to foster scalability by separating microservices.**
2) Based on the provides and demands relations, the ßMACH protocol describes knowledge propagation. We already mentioned the propagation for the aspects without active management. The knowledge propagation for active management is interesting for a software engineer. How is the knowledge transferred and converted, and which knowledge is it? In Figure 3, the starting point is the concerns. Individual microservices handle the system's business concerns (row outside responsibilities). The microservice team's responsibilities are the basis for the outside interfaces and the outside dependencies. Thus, the product properties are indirectly based on the concerns. In other words, the business concerns of the microservice need to be defined first. Afterward, the microservice team cares about creating and operating the microservice. **The team cares for the microservice. The ßMACH protocol points out that the team needs a defined business concern as a starting point and then manages the service creation and operation based on the concerns. Another influence on the team is not present.** (See Section 4 for other management decisions and strategies.)
3) Only one kind of description for roles is present in the ßMACH protocol (Fig. 3). The roles are not exactly defined. In other examples of ßMACH protocols, we have seen concrete roles like software developers, architects, and designers. In Figure 3, there is a DevOps team. The roles this team needs are not fully defined. Based on the understanding of DevOps, the roles are reasonable to perform the given tasks. **The ßMACH protocol does not point out that the DevOps team needs to adapt to the microservice's business concern. For a concrete project, we need to define and instantiate the abstract definition of roles.**
4) The process knowledge is given by self-management of the DevOps team. Similar to the roles, this is not concrete. Nevertheless, DevOps is the idea of small teams and self-management. A concrete team should give more details on how to perform self-management. **ßMACH points out the DevOps team's self-**

managed process. Thus, inadequate influences on the team have to be omitted. It is also a consequence of the independence of microservices.

5) The definition of the management concept in ßMACH (Fig. 1) holds four easy-to-read bullet points. It is very minimal, easy to remember, fast to understand, and interpretation is present and referrable at any time (Fig. 3). Based on our observation, it is very helpful to have an explicit management process. It makes the process easier, reduces conflicts, and enables improvements. Also, a change in the process gets obvious, and changes can be explicitly discussed. **The ßMACH protocol is compact, and it is easy to understand the definition of the management process. So, the planned process is written down and can be referred to later on.** (See Section 4 for changing the management strategy and updating the protocol.)
6) The effort to create a ßMACH protocol is not very high and no special knowledge or skills are needed. To describe microservices, you need to understand microservices. So, you can create a protocol in about two hours on a whiteboard with the DevOps team. Afterward, the process is clear to all team members. We have also discussed two weeks about a single ßMACH protocol. We discussed the management process, and we learned a lot. Based on filling the ßMACH protocol, we identified the gaps in the process, found borderline cases, and nailed down the differences between our idea of the process and the practical doing. At least based on our observation, the ßMACH method supported us. **A ßMACH protocol is created in some hours and can help to improve the management process.**
7) ßMACH is a communication helper. The terms in the protocol support understandability. The description of the process by one person is easier to understand for the team. It was even possible to identify misunderstandings between persons. If two persons answered a management aspect in the ßMACH protocol differently, the process was not yet clear. **The definition of terms and the systematics of the ßMACH protocol support the communication of the involved persons and avoid misunderstandings in the management process.**
8) Our definition of microservices is strong for explaining the essentials of the concept. To allow minimal dependencies is a concession to practical implementations of microservice systems. Nevertheless, isolation is not easy to realize. Especially for transforming legacy systems to microservices, the definition gives a goal, not the transformation or an intermediate step. Thus, a practical realization of a microservice system probably sacrifices strict isolation and decides to manage the consequences instead of dealing with the realization of strict isolation. **We use a microservice definition to point out the advantages of strict isolation. A real-world system uses potbelly less strict definitions with reduced isolation. Especially, the transformation of a monolith into a microservice system will not hold our definition. In such a case, the ßMACH protocol will look different.**

7. Conclusion

Based on performing the ßMACH method, we can state two kinds of findings.

First, the ßMACH method is helpful for the analysis of software management processes and supports the management. The method is systematic and defines terms to describe the process. So, it supports analyses, such as the understanding of the process by the development team and learning the process by all team members. Also, the method is easy and fast to perform and thus efficient.

Second, the isolation of individual microservices supports the development team. The team can avoid many aspects of management. In addition, the team can perform all the knowledge representation and transformation to develop an individual microservice. There are no supplementary relations or dependencies to the team. So, the number of teams can be scaled without overhead to individual teams.

We close this paper by introducing you to fill a ßMACH protocol for your software management process and take value from the method. To learn more about your management proceeding and how your colleagues understand it.

Acknowledgements

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 414984028 – SFB 1404 FONDA

References

[1] 4Soft GmbH in Zusammenarbeit mit dem Informationstechnikzentrum Bund: *V-Modell XT Bund, Das Referenzmodell für Systementwicklungsprojekte in der Bundesverwaltung*, Informationstechnikzentrum Bund im Auftrag des Beauftragten der Bundesregierung für die Informationstechnik, version: 2.3 ed., 2019.

[2] Alagarasan V.: Seven Microservices Anti-patterns, 2015. www.infoq.com/articles/seven-uservices-antipatterns.

[3] Anderson D., Anderson L.: *Beyond Change Management: How to Achieve Breakthrough Results Through Conscious Change Leadership*, John Wiley & Sons, 2010.

[4] Armstrong D.J.: The quarks of object-oriented development, *Communications of the ACM*, vol. 49(2), pp. 123–128, 2006. doi: 10.1145/1113034.1113040.

[5] Assadi A.: What are microservices?, 2016. https://www.linkedin.com/pulse/what-microservices-abtin-assadi/.

[6] Basili V.R.: Software modeling and measurement: the Goal/Question/Metric paradigm, Technical report, 1992.

[7] Berges M.P.: *Object-Oriented Programming through the Lens of Computer Science Education*, Dissertation, Technische Universität München, München, 2015.

[8] Bernard P.: *Foundations of ITIL® 2011 Edition*, Van Haren, 2011.

[9] Bourque P., Fairley R.E. (eds.): *Guide to the Software Engineering Body of Knowledge. Version 3. SWEBOK®. A project of the IEEE Computer Scociety*, IEEE Computer Society Press, Washington, DC, USA, 2014.

[10] Brewster E., Griffiths R., Lawes A., Sansbury J.: *IT Service Management: A Guide for ITIL Foundation Exam Candidates*, BCS, The Chartered Institute for IT, 2012.

[11] Bucchiarone A., Dragoni N., Dustdar S., Lago P., Mazzara M., Rivera V., Sadovykh A. (eds.): *Microservices – Science and Engineering*, Springer, Cham, 2020. doi: 10.1007/978-3-030-31646-4.

[12] Bungay S.: *The Art of Action: How Leaders Close the Gaps between Pans, Actions and Results*, Nicholas Brealey, 2010.

[13] Cater-Steel A., Tan W.G.: itSMF Australia 2005 Conference: Summary of ITIL Adoption Survey Responses, Technical report, University of Southern Queensland, 2005. https://research.usq.edu.au/item/9y541/itsmf-australia-2005-conference-summary-of-itil-adoption-survey-responses.

[14] Cechini F., Ice R., Binkley D.: Systems Engineering Guidebook for Intelligent Transportation Systems, Technical report, U.S. Department of Transportation, Federal Highway Administration, California Division, 2009.

[15] Clarke P., Mesquida A.L., Ekert D., Ekstrom J.J., Gornostaja T., Jovanovic M., Johansen J., *et al.*: An Investigation of Software Development Process Terminology. In: P. Clarke, R.V. O'Connor, T. Rout, A. Dorling (eds.), *Software Process Improvement and Capability Determination. SPICE 2016. Communications in Computer and Information Science*, pp. 351–361, Springer, Cham, 2016. doi: 10.1007/978-3-319-38980-6_25.

[16] Copei S., Zündorf A.: How to Synchronize Microservices. In: *Proceedings of International Conference on Microservices*, University of Applied Sciences and Arts Dortmund, Germany, 2019. https://www.conf-micro.services/2019/papers/Microservices_2019_paper_16.pdf.

[17] Crnkovic I., Sentilles S., Vulgarakis A., Chaudron M.R.V.: A Classification Framework for Software Component Models, *IEEE Transactions on Software Engineering*, vol. 37(5), pp. 593–615, 2011. doi: 10.1109/tse.2010.83.

[18] Dennis A., Wixom B., Tegarden D.: *Systems Analysis and Design with UML*, Wiley, 2009.

[19] Dragoni N., Giallorenzo S., Lafuente A.L., Mazzara M., Montesi F., Mustafin R., Safina L.: Microservices: Yesterday, Today, and Tomorrow. In: M. Mazzara, B. Meyer (eds.), *Present and Ulterior Software Engineering*, pp. 195–216, Springer, Cham, 2017. doi: 10.1007/978-3-319-67425-4_12.

[20] Duell M., Goodsen J., Rising L.: Non-software examples of software design patterns. In: *OOPSLA '97: Addendum to the 1997 ACM SIGPLAN conference on Object-oriented programming, systems, languages, and applications (Addendum)*, pp. 120–124, 1997. doi: 10.1145/274567.274592.

[21] Dyck A., Penners R., Lichter H.: Towards Definitions for Release Engineering and DevOps. In: *2015 IEEE/ACM 3rd International Workshop on Release Engineering*, pp. 3–3, 2015. doi: 10.1109/RELENG.2015.10.

[22] Edwards D.W.: *Out of the Crisis*, 1986. doi: 10.7551/mitpress/11457.001.0001.

[23] Elliott G.: *Global Business Information Technology: An Integrated Systems Approach*, Pearson Addison Wesley, 2004. https://books.google.ru/books?id=qGfzMlfgzEcC.

[24] Everett G.D., Raymond McLeod J.: *Software Testing; Testing Across the Entire Software Development Life Cycle*, John Wiley & Sons, Ltd, 2006. doi: 10.1002/9780470146354.fmatter.

[25] Fernández D.M., Wagner S.: Naming the pain in requirements engineering: A design for a global family of surveys and first results from Germany, *Information and Software Technology*, vol. 57, pp. 616–643, 2016. doi: 10.1016/j.infsof.2014.05.008.

[26] Fowler M.: *Patterns of Enterprise Application Architecture*, Addison-Wesley, 2003.

[27] Fowler M.: *Domain-specific languages*, Pearson Education, 2010.

[28] Fritzsch J., Bogner J., Wagner S., Zimmermann A.: Microservices in the German Industry: Insights into Technologies, Characteristics, and Software Quality. In: *Proceedings of International Conference on Microservices*, University of Applied Sciences and Arts Dortmund, Germany, 2019. https://www.conf-micro.services/2019/papers/Microservices_2019_paper_25.pdf.

[29] Gabbrielli M., Lanese I., Zingaro S.P.: Microservice-Oriented Computing for the Internet of Things. In: *Proceedings of International Conference on Microservices*, University of Applied Sciences and Arts Dortmund, Germany, 2019. https://www.conf-micro.services/2019/papers/Microservices_2019_paper_3.pdf.

[30] Ginger L.: Embrace and Exploit Change as a Program Manager: Guidelines for Success, *Journal of Change Management*, vol. 10, pp. 2–13, 2012.

[31] Hasselbring W.: Software Architecture: Past, Present, Future. In: V. Gruhn, R. Striemer (eds.), *The Essence of Software Engineering*, pp. 169–184, Springer International Publishing, Cham, 2018. doi: 10.1007/978-3-319-73897-0_10.

[32] Hiatt J.M.: *ADKAR: a model for change in business, government, and our community*, Prosci, 2006.

[33] Hilbrich M., Bountris V.: Are Workflows a Language to Solve Software Management Challenges? – A ßMACH Based Analysis. In: *New Trends in Intelligent Software Methodologies, Tools and Techniques*, Frontiers in Artificial Intelligence and Applications, vol. 355, pp. 221–232, IOS Press, 2022. doi: 10.3233/FAIA220253.

[34] Hilbrich M., Frank M.: Abstract Fog in the Bottle – Trends of Computing in History and Future. In: *2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, pp. 519–522, 2018. doi: 10.1109/SEAA.2018.00089.

[35] Hilbrich M., Jakobs C., Werner M.: Do Microservices Prevent High Qualitative Code? In: *Proceedings of International Conference on Microservices*, University of Applied Sciences and Arts Dortmund, Germany, 2019. https://www.conf-micro.services/2019/papers/Microservices_2019_paper_21.pdf.

[36] Hilbrich M., Lehmann F.: Discussing Microservices: Definitions, Pitfalls, and their Relations. In: *2022 IEEE International Conference on Services Computing (SCC)*, pp. 39–44, IEEE Computer Society, Los Alamitos, CA, USA, 2022. doi: 10.1109/SCC55611.2022.00019.

[37] Hilbrich M., Lehmann F.: ßMACH – A Software Management Guidance. In: D.G. Reichelt, R. Müller, S. Becker, W. Hasselbring, A. van Hoorn, S. Kounev, A. Koziolek, R. Reussner (eds.), *Symposium on Software Performance 2021*, CEUR-WS, 2022. http://ceur-ws.org/Vol-3043/.

[38] Hui A.: Lean Change: Enabling Agile Transformation through Lean Startup, Kotter and Kanban: An Experience Report. In: *2013 Agile Conference*, pp. 169–174, 2013. doi: 10.1109/agile.2013.22.

[39] IEEE Standard for Developing a Software Project Life Cycle Process, *IEEE Std 1074-2006 (Revision of IEEE Std 1074-1997)*, pp. 1–110, 2006. doi: 10.1109/IEEESTD.2006.219190.

[40] ISO/IEC 15504-1:2004 Information technology – Process assessment – Part 1: Concepts and vocabulary, 2004. https://www.iso.org/obp/ui/#iso:std:iso-iec:15504:-1:ed-1:v1:en.

[41] ISO/IEC TS 24748-1:2016 Systems and software engineering – Life cycle management – Part 1: Guidelines for life cycle management, 2016. https://www.iso.org/obp/ui/#iso:std:iso-iec:ts:24748:-1:ed-1:v1:en.

[42] ISO/IEC/IEEE 12207:2017 Systems and software engineering – Software life cycle processes, 2017. https://www.iso.org/obp/ui/en/#iso:std:iso-iec-ieee:12207:ed-1:v1:en.

[43] Jabbari R., bin Ali N., Petersen K., Tanveer B.: What is DevOps? A Systematic Mapping Study on Definitions and Practices. In: *XP '16 Workshops: Proceedings of the Scientific Workshop Proceedings of XP2016*, Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2962695.2962707.

[44] Johnson P., Ekstedt M., Jacobson I.: Where's the Theory for Software Engineering?, *IEEE Software*, vol. 29(5), 96, 2012. doi: 10.1109/MS.2012.127.

[45] Kroll P., Kruchten P.: *The Rational Unified Process Made Easy: A Practitioner's Guide to the RUP*, Addison-Wesley Professional, 2003.

[46] Lange M., Koschel A., Hausotter A.: Microservices in Higher Education – Migrating a Legacy Insurance Core Application. In: *Proceedings of International Conference on Microservices*, University of Applied Sciences and Arts Dortmund, Germany, 2019. https://www.conf-micro.services/2019/papers/Microservices_2019_paper_8.pdf.

[47] Lau K.K., Wang Z.: Software Component Models, *IEEE Transactions on Software Engineering*, vol. 33(10), pp. 709–724, 2007. doi: 10.1109/tse.2007.70726.

[48] Lea G.: Why "Don't Use Shared Libraries in Microservices" is Bad Advice, 2016. http://www.grahamlea.com/2016/04/shared-libraries-in-microservices-bad-advice/.

[49] Lewis J., Fowler M.: Microservices: a definition of this new architectural term, 2014. http://martinfowler.com/articles/microservices.html.

[50] Linthicum D.: Chapter 1: Service Oriented Architecture (SOA), 2016. https://web.archive.org/web/20160206132542/https://msdn.microsoft.com/en-us/library/bb833022.aspx#. Accessed 18.06.2020.

[51] Lu N., Glatz G., Peuser D.: Moving mountains – practical approaches for moving monolithic applications to Microservices. In: *International Conference on Microservices*, University of Applied Sciences and Arts Dortmund, Germany, 2019. https://www.conf-micro.services/2019/papers/Microservices_2019_paper_30.pdf.

[52] Marrone M., Kolbe L.M.: Impact of IT Service Management Frameworks on the IT Organization. An Empirical Study on Benefits, Challenges, and Processes, *Business & Information Systems Engineering*, vol. 3(1), pp. 5–18, 2011. doi: 10.1007/s12599-010-0141-5.

[53] Martin R.C.: *Clean Code: A Handbook of Agile Software Craftsmanship*, Prentice Hall, Upper Saddle River, NJ, 2008. https://www.safaribooksonline.com/library/view/clean-code/9780136083238/.

[54] Maschio B.: Updating the current Jolie microservices based Document Management System to include electronic invoicing. In: *International Conference on Microservices*, University of Applied Sciences and Arts Dortmund, Germany, 2019. https://www.conf-micro.services/2019/papers/Microservices_2019_paper_15.pdf.

[55] Mauro T.: Adopting Microservices at Netflix: Lessons for Architectural Design, 2015. https://www.nginx.com/blog/microservices-at-netflix-architectural-best-practices/.

[56] Microsoft: What are microservices?, 2022. https://azure.microsoft.com/en-us/solutions/microservice-applications/. Accessed 30.08.2020.

[57] Microsoft: Gründe für einen Microservice-Ansatz zum Erstellen von Anwendungen, Version: Jun 14, 2019. https://docs.microsoft.com/de-de/azure/service-fabric/service-fabric-overview-microservices.

[58] Nadareishvili I., Mitra R., McLarty M., Amundsen M.: *Microservice Architecture. Aligning principles, practices, and culture*, O'Reilly Media, 2016.

[59] Nash J., Ehrenfeld J.: Code Green: Business Adopts Voluntary Environmental Standards, *Environment: Science and Policy for Sustainable Development*, vol. 38(1), pp. 16–45, 1996. doi: 10.1080/00139157.1996.9930973.

[60] Naur P., Randell B.: Software Engineering: Report of a conference sponsored by the NATO Science Committee, Garmisch, Germany, 7–11 Oct. 1968, Brussels, Scientific Affairs Division, NATO. 1969.

[61] Newman S.: *Building Microservices*, O'Reilly Media, 2015.

[62] Object Management Group (OMG): Automated Function Points (AFP), https://www.omg.org/spec/AFP/1.0/PDF, 2014. Accessed 17.08.2020.

[63] Popper K.R.: *Objective Knowledge: An Evolutionary Approach*, Oxford, England: Oxford University Press, 1972.

[64] Popper K.R., Eccles J.C.: *The Self and its Brain: An Argument for Interactionism*, Springer, 1977.

[65] Pratt M.: Microservice Pitfalls & AntiPatterns, Part 1, 2016. https://homeadvisor.tech/software-antipatterns-microservices/.

[66] Ralph P.: Toward Methodological Guidelines for Process Theories and Taxonomies in Software Engineering, *IEEE Transactions on Software Engineering*, vol. 45(7), pp. 712–735, 2019. doi: 10.1109/TSE.2018.2796554.

[67] Rational Software: Rational Unified Process, Best Practices for Software Development Teams, 1998. https://www.ibm.com/developerworks/rational/library/content/03July/1000/1251/1251_bestpractices_TP026B.pdf. Accessed 18.06.2020.

[68] Richards M.: *Microservices Antipatterns and Pitfalls*, O'Reilly Media, 2016.

[69] Scacchi W.: Process models in software engineering, *Encyclopedia of Software Engineering*, 2001. doi: 10.1002/0471028959.sof250.

[70] Schwaber K., Sutherland J.: *The Scrum Guide. The Definitive Guide to Scrum: The Rules of the Game*, 2020. https://scrumguides.org/docs/scrumguide/v2020/2020-Scrum-Guide-US.pdf.

[71] Simon H.A.: *The Sciences of the Artificial*, MIT Press, 1996. doi: 10.7551/mitpress/12107.001.0001.

[72] Sjøberg D.I.K., Dybå T., Anda B.C.D., Hannay J.E.: *Building Theories in Software Engineering*, pp. 312–336, Springer London, London, 2008. doi: 10.1007/978-1-84800-044-5_12.

[73] Sneed H.M.: *Software Management*, Verlagsgesellschaft Rudolf Müller, Köln, 1987.

[74] Sommerville I.: *Software Engineering,* 10th edition, Pearson Education Limited, Boston, 2016.

[75] Staples M.: Critical rationalism and engineering: ontology, *Synthese*, vol. 191(10), pp. 2255–2279, 2014. doi: 10.1007/s11229-014-0396-3.

[76] Staples M.: Critical rationalism and engineering: methodology, *Synthese*, vol. 192(1), pp. 337–362, 2015. doi: 10.1007/s11229-014-0571-6.

[77] Stein A., Zillekens M., Khan M.: A Microservice architecture for monitoring, processing and predicting climate data in animal husbandry. In: *International Conference on Microservices*, University of Applied Sciences and Arts Dortmund, Germany, 2019. https://www.conf-micro.services/2019/papers/Microservices_2019_paper_28.pdf.

[78] Steinacker G.: Why Microservices?, 2016. https://www.otto.de/jobs/en/technology/techblog/blogpost/why-microservices_2016-03-20.php. Accessed 08.2020.

[79] Szyperski C., Gruntz D., Murer S.: *Component Software: Beyond Object-Oriented Programming*, Addison-Wesley Educational Publishers Inc, 2002.

[80] The Standish Group International, Inc.: The CHAOS Report (1994). Technical report, 1994.

[81] The Standish Group International, Inc.: Chaos Report 2015. Technical report, 2015.

[82] Thompson M.: Why SaaS and Microservices are Critical to Developing in the Cloud, 2015. https://www.rightbrainnetworks.com/2015/01/29/why-saas-and-microservices-are-critical-to-developing-in-the-cloud/.

[83] Tilkov S.: Microservices: A Taxonomy. In: *International Conference on Microservices*, University of Applied Sciences and Arts Dortmund, Germany, 2019. https://www.conf-micro.services/2019/papers/Microservices_2019_paper_29.pdf.

[84] Tracz W.: DSSA (Domain-Specific Software Architecture): Pedagogical Example, *SIGSOFT Software Engineering Notes*, vol. 20(3), pp. 49–62, 1995. doi: 10.1145/219308.219318.

[85] Ullenboom C.: *Java ist auch eine Insel*, Galileo Computing, Bonn, 6., aktualisierte und erweiterte Auflage ed., 2007. https://openbook.rheinwerk-verlag.de/javainsel/.

[86] U.S. Environmental Protection Agency, Office of Atmospheric Programs, Climate Protection Partnerships Division: Clean Energy-Environment Guide to Action, Policies, Best Practices, and Action Steps for States, 2006. https://web.archive.org/web/20120713125427/http://www.epa.gov/statelocalclimate/documents/pdf/guide_action_full.pdf. Accessed 13.08.2020.

[87] Vasanthapriyan S., Tian J., Xiang J.: A Survey on Knowledge Management in Software Engineering. In: *2015 IEEE International Conference on Software Quality, Reliability and Security – Companion*, pp. 237–244, 2015. doi: 10.1109/QRS-C.2015.48.

[88] Vega N.: Answering Your Microservices Webinar Questions, 2015. https://www.ibm.com/blogs/bluemix/2015/02/answering-microservices-webinar-questions/#q1.

[89] Wetherill J.: Microservices and PaaS (Part I), 2014. https://dzone.com/articles/microservices-and-paas-part-1.

[90] Wilde N., Gonen B., El-Sheikh E., Zimmermann A.: Approaches to the Evolution of SOA Systems. In: E. El-Sheikh, A. Zimmermann, L.C. Jain (eds.), *Emerging Trends in the Evolution of Service-Oriented and Enterprise Architectures*, pp. 5–21, Springer International Publishing, Cham, 2016. doi: 10.1007/978-3-319-40564-3_2.

[91] Wohlin C., mite D., Moe N.B.: A general theory of software engineering: Balancing human, social and organizational capitals, *Journal of Systems and Software*, vol. 109, pp. 229–242, 2015. doi: 10.1016/j.jss.2015.08.009.

[92] Wolff E.: *Microservices: Grundlagen flexibler Softwarearchitekturen*, dpunkt.verlag GmbH, Heidelberg, 2015.

[93] Wolff E.: Why Microservices Fail: An Experience Report. In: *International Conference on Microservices*, University of Applied Sciences and Arts Dortmund, Germany, 2019. https://www.conf-micro.services/2019/papers/Microservices_2019_paper_18.pdf.

Appendix

ßMACH Protocol to Evaluate the Management of Software Artifacts

This is the ßMACH protocol. This protocol aims to improve the software artifact management. The target audience is persons in software development, software operation, software mentions, corresponding team leaders, managers of software artifacts, and those interested in improving the management of software artifacts on a small or big scale.
The method is to give the definition of software artifact management in this protocol and to give descriptions of different aspects of software management. To support all management aspects, you provide all descriptions (fill in all fields of the table below) based on your management definition. By giving the descriptions and extending the management definition until you have every description, you improve your management definition to respect all the different aspects of management given in this protocol. Your definition of the management process is not allowed to have parts not used by a description, So you avoid unnecessary parts of the definition.
You fill in this protocol before starting the software management process, ideally. So, you have the definition of the process present in time. During the management process, you can refer to it all the time. This protocol is an explicit description of your management concept. If something changes, you can create a new version of this protocol. From time to time, you check this protocol retrospective and identify which descriptions and parts of the definition are helpful and which aspects need improvement for the next version of this protocol or another protocol. So, you can improve your management skills.
The structure of the ßMACH protocol is as follows: A separate box asks for the context information. The protocol starts with the work package responsibilities to check your principal tasks. Following is the definition of your management concept and the table with the descriptions (for the different management aspects) you have to give. ßMACH recommends starting with the context, selecting the work packages, and then developing the definition while providing the descriptions stepwise.

Work Package Responsibilities:

Check this item if your team is responsible for product development. It means your team directly influences the creation of software artifacts, e.g., by programming modeling or supervising. Additionally, your team finishes the development.

☑ Finalizing Product Development

Check this item if your team is responsible for product maintenance. It means your team directly influences (non-functional) changes to the software artifacts while the artifacts are in operation, e.g., you are in a DevOps team or have to provide a bug fix. Additionally, the maintenance is not moved to another team later on.

☐ Finalizing Product Maintenance

Check this item if your team is responsible for product improvement. It means your team directly influences (functional) changes to software artifacts, e.g., you add additional features or develop an extended version. Additionally, no other team is improving the software artifacts later on.

☐ Finalizing Product Improvement

Definition of the Management Concept:

Here, you provide the definition of the software artifact management concept. You have to use each part of the definition in the description part below. Also, you have to base all descriptions (see below) on parts of this definition. The definition has to mark individual parts and give a reference for each part. You can use, e.g., an enumeration.
As a result, the definition is complete because you base all descriptions on the next page on the definition, and it is minimal because it does not have unused and unnecessary parts.

1. The software is created by a magical box.
2. The magical box works fully automatic.
3. The software is used exactly once, by the team itself.
4. Our team has to ask other teams to get the magical box and we have to return it.
5. The persons in our team can do tasks that not demand special knowledge or skills.
6. We do not want to learn from the project and we get a fixed price, so we deny documentation tasks.

Figure A1. Definition of software development toy example in the ßMACH protocol. The definition gives the numbers of the bullet points/parts. (See Fig. A3 and https://doi.org/10.5281/zenodo.10992007 for more details.)

Context (User/Team/Context Information):

Give the name of the person or persons that fill in this ßMACH protocol. Add company name, department, or whatever is needed to identify the person later.

Name of the Filler: Marcus Hilbrich

Give the team managing the software artifacts. It can be a single person managing some product artifacts, a five-person development team developing a single product component, or a whole organization managing multiple products. The definition of the team is essential because it gives the context for the descriptions in the later part and other questions. So, it highly influences the answers. You describe the team understandably, e.g., name all members or a generally known team name. In addition, the represented team defines the internals, and the cooperating teams define the external border.

Represented Team (Internal Border): Team 002

Give the teams you have to or can cooperate with. It can be teams of your organization you can emerge resources with, teams of other organizations that work on the same overall product, and teams you have to deliver or get something from (e.g., the billing team or a requirements engineering team). The set of teams defines what you can influence indirectly. The difference between cooperating teams and externals is that you cannot cooperate with externals (you have no direct influence on them). It is the outside border.

Cooperating Teams (External Border): All teams of the Magical Box Organisation

Give the artifacts to manage. It can be a set of concrete artifacts, a product, a project, or another understandable description. The description of the artifacts is essential because it describes what to manage, where there are possibly interfaces to artifacts that are not managed by your team or cooperating teams, etc.

Managed Artifacts: The boring game No. 3

Give the calendar date or timespan for filling this document. So, the coordination with other documents is possible.

Date: 04/29/2020

The version of the filled ßMACH protocol. The version number identifies minor or major changes and distinguishes documents.

Version of Document: 0.1.0 Initial Version

Here, you can put everything that has no other place in the document. It can be general remarks, changes to the last version, open questions, problems with the document itself, or whatever you can think of.

Comment:

This document is the description of Questionaries. All elements of the questionary are given in black, explanations to the elements are given in grey (next to the described element, and a very basic example of filling the ßMACH protocol is given in blue.)
Background colors in the following table are examples and not general, see the description of the table for future information.

Figure A2. Context or meta-data of software development toy example in the ßMACH protocol. (See Fig. A3 and https://doi.org/10.5281/zenodo.10992007 for more details.)

In the following table, you have to give the descriptions. You have to describe the management of different aspects based on the definition of the management concept (see above). The descriptions always concern the user/team/context information (see above). So, you always answer for your team and the artifacts you have given.
In case of multiple exemplifications that are very similar and based on the same part of the definition, give all corresponding fields of the table a unique text background color (different from white).
Right, in the field of an explanation, you have to give the set of parts of the definition of the management concept (an understandable reference) that you use for the corresponding description.
If nothing needs to be managed, based on the definition, mark the right part of the field (the references to the definition of the management concept) with a (light) green background color. An example is when you define the field as not relevant.
If the field is provided by another one and thus nothing needs to be managed, mark the right part of the field (the references to the definition of the management concept) with a dark green background color. An example is when a role provides knowledge, so the knowledge probably does not need individual management.
If the field is demanded by another one, mark the right part of the field (the references to the definition of the management concept) with a violet background color. (When the background is already green or light green, add a violate spot instead.) An example is when a role is needed to generate knowledge in a form defined by another field.

Figure A3. The text is part of the hints to the ßMACH protocol. It describes the filling in of the key aspects. The key aspects are separated into Fig. A4, A5, and A6. ßMACH describes a set of key aspects. Each cell of the table (Fig. A4, A5, and A6) represents an aspect. The right part of each cell holds the references to the definition in the ßMACH protocol (Fig. A1). ßMACH defines coloring. Based on the management process, we use light green in the right part of a cell for aspects that do not need active management. Active management means that an aspect is realized without a need for action. The darker green indicates that an aspect is also performed without needing active management but is provided by another. We use violet for aspects, used or required by additional ones. Such an aspect indicates a special interest. Arrows with a peak-end describe a provides-relation. The aspect on the peak is provided by the other. A round end arrow gives a demand relation. The other aspect needs the one at the rounded end. The left part of the call can be colored, too. If the left part of multiple cells uses the same color, the cell's descriptions are equal or very similar. The described aspects in this figure are all based on the other parts of the ßMACH protocol, provided in Fig. A1 and A2. Download the ßMACH protocol as PDF to zoom in and explore the protocol (https://doi.org/10.5281/zenodo.10992007)

	Product Knowledge		Demanded Knowledge		Roles		Process Knowledge		Process Information	
Explanation for Aspects the Team is Not Responsible for:										
Product Development Cross the fields in this row when your work package responsibility covers product development (check above).	If not crossed, give the information and knowledge your team has to provide to the development team(s). What are the demands for the software artifacts to build, what are the requirements, whom to contact, who has additional knowledge, who is responsible, who provides changes to consider, etc.? ———	–	If not crossed, give to be learned or not yet present knowledge that is important for the development team. Is domain knowledge required, are the software artifacts part of a specific process that has to be known, is knowledge about needed tools demanded, is knowledge for understanding provided documents needed, and similar? ———	–	If not crossed, give the required roles in your team to provide information and knowledge in this row, e.g., by teaching, learning, or using artifacts. ———	–	If not crossed, give the process your team uses to provide information and knowledge in this row. What is the (kind of) artifacts you create to provide information and knowledge? How are the artifacts created? What are the single steps, intermediated artifacts, the process, etc? ———	–	If not crossed, give information to collect during activities triggered by this row. Is it needed to write working hours, provide a list of invoked persons, have an access control list, is a protocol of tests required for certification, is a process documentation needed by an authority, or similar? ———	–
Product Maintenance Cross the fields in this row when your work package responsibility covers product maintenance (check above).	If not crossed, give the information and knowledge your team has to provide for the maintenance team. What is the software architecture, which compilers have to be used, where to find documentation, what are the requirements to fulfill, how to deploy the software, how to operate the software, etc? Based on 3, no maintains is needed.	3	If not crossed, give to be learned or not yet present knowledge that is important for the maintenance team. Is domain knowledge required? Are the software artifacts part of a specific process that has to be known? Is knowledge about used tools needed? Does the understanding of the provided documents require additional knowledge? Or other questions. Based on 3, no maintains is needed.	3	If not crossed, give the needed roles in your team to provide information and knowledge in this row, e.g., by teaching, learning, or using artifacts. Based on 3, no maintains is needed.	3	If not crossed, give the process your team uses to provide information and knowledge in this row. What is the (kind of) artifacts your team creates to provide information and knowledge? How are the artifacts created, what are the single steps, intermediated artifacts, the process, etc.? Based on 3, no maintains is needed.	3	If not crossed, give information collected during activities triggered by this row. Is it needed to write working hours, provide a list of invoked persons, have an access control list, is a test protocol required for certification, is a process documentation needed by an authority, or similar? Based on 3, no maintains is needed.	3
Product Improvement Cross the fields in this row if your work package responsibility covers product improvement (check above).	If not crossed, give the information and knowledge your team has to provide for the improvement team. What is the software architecture, which compilers have to be used, where to find documentation, what are the original requirements to fulfill, how to deploy the software, how to operate the software, etc? Based on 3, no improvement is needed.	3	If not crossed, give to be learned or not yet present knowledge that is important for the improvement team. Is domain knowledge required? Are the software artifacts part of a specific process that has to be known? Is knowledge about used tools needed? Does the understanding of the provided documents require additional knowledge? Or other questions. Based on 3, no improvement is needed.	3	If not crossed, give the roles in your team to provide information and knowledge in this row, e.g., by teaching, learning, or using artifacts. Based on 3, no maintains is needed.	3	If not crossed, give the process your team uses to provide information and knowledge in this row. What are the (kind of) artifacts your team creates to provide information and knowledge? How are the artifacts created, what are the single steps, intermediated artifacts, the process, etc.? Based on 3, no maintains is needed.	3	If not crossed, give information to collect during activities triggered by this row. Is it needed to write working hours, provide a list of invoked persons, have an access control list, is a test protocol required for certification, is a process documentation needed by an authority, or similar? Based on 3, no maintains is needed.	3

Figure A4. Description of software development toy example based on the ßMACH protocol. Presented is a set of key aspects. (See Fig. A3 and https://doi.org/10.5281/zenodo.10992007 for more details.)

	Product Knowledge		Demanded Knowledge		Roles		Process Knowledge		Process Information	
Explanation for Aspects the Team is Responsible for:										
Inside Product Properties You have to give the management of the properties of the artifacts your team has to manage. It can be all kinds of requirements for the artifacts. In this row, you consider the product properties directly related to your team and demands by cooperating teams (related to the internal border).	Give the knowledge about the internal product properties. You describe the kind of used artifacts, the (kind of) knowledge about the product you need to manage, etc. Examples are requirement lists, descriptions of algorithms you implement, and domain knowledge. Based on 2 the box holdes the knowledge, a cooperation to other teams is not needed.	2	Give the knowledge your team needs (but does not yet have) to realize artifact management. Do you need knowledge, e.g., to use tools, communicate with other teams, or explore solutions? To get the box (4) or use it (2), no additional knowledge is needed.	2, 4	Give the roles you need in your team to handle product properties, e.g., by providing knowledge or performing tasks. To ask the other teams about the box does not demand a specific role.	4, 5	Give needed knowledge of your team to perform the process to realize the product properties. It can be (a part of) a software process model, a management concept, outsourcing, or whatever you want to do. Based on 1, 2 and 4 we have to get the magical box and start it, so no specific knowledge is needed (the person knows what to do).	1, 2, 4	Give the information that has to be collected while performing the management process in this row. Is it demanded to collect test results, do you have to give information for billing, is version management mandated, do you have to rate the used process, or something else? Not demanded by the definition.	6
Outside Product Properties You have to give the management of the properties of the artifacts your team (together with the cooperating teams) has to manage. It can be all kinds of requirements for the artifacts. In this row, you consider the product properties demanded by external parties (not by cooperating teams). It is related to the external border.	Give the knowledge about the external product properties. You describe the kind of artifacts to use, the (kind of) knowledge regarding the product you manage, etc. Examples are requirement lists, descriptions of the algorithm to implement, and domain knowledge. The box holdes the knowledge, the team does not managed it, additional knowledge is not needed.	2	Give the knowledge your team needs (but does not yet have) to realize artifact management. Does your team need the knowledge to use tools, communicate with other teams, explore solutions, etc.? To use the box (2), no additional knowledge is needed.	2	Give the roles in your team to handle product properties, e.g., by providing knowledge or performing tasks. To start the magical box does not demand a specific role.	2, 5	Give the knowledge needed to perform the process that realizes the product properties. It can be (a part of) a software process model, a management concept, outsourcing, or whatever you want to do. Based on 2, no noticeable process is needed.	2	Give the information to collect while performing the management process in this row. Is it demanded to collect test results, to give billing information, is version management mandated, do you have to rate the used process, or something else? Not demanded by the definition.	6
Inside Dependencies You have to give the management of the dependencies your team has to manage. It can be all kinds of dependencies of/to the artifacts, dependencies of/to persons, dependencies to tooling, and similar. In this row, you consider dependencies to cooperating teams (related to the internal border).	Give the (kind of) knowledge and the representation of the knowledge needed for your artifacts that your team delivers or shares with cooperating teams and your team demands. It can be cooperative development of artifacts like libraries, the demand to get requirements for your artifacts, etc. The magical box holds the knowledge, we depend on the other teams to get the box.	2, 4	Give the (kind of) knowledge and the representation of the knowledge needed for your team to work but which is not (completely) present in your team. Is it planned to teach your team the usage of tools, is it demanded to transfer the domain knowledge to your team (by teaching or adding a member), etc? The description does not lists knowledge not yet part of the team that is needed.	2	Give the roles in your team that handle dependencies in this row. Give the role that has to learn knowledge or use a shared artifact. No knowledge to manage is given in this row, so no role is needed.		Give dependencies in executing the management process. Is the ongoing of your team managed by another team (e.g., the Scrum master or product owner is not part of your team)? No knowledge that is needed to processed is given in this row, so no management process is needed.		Give dependencies of process information in this row. For example, if your team collects information when other teams deliver knowledge or how much your team is billed by cooperating teams. Not demanded by the definition.	6
Outside Dependencies Give the management of the dependencies your team (together with the cooperating teams) has to manage. It can be all kinds of dependencies of/to the artifacts, of/to persons, to tooling, and similar. In this row, you have to consider the dependencies demanded by external (not by cooperating teams) parties (related to the external border). External demands are generally risky.	Give the (kind of) knowledge and the representation of the knowledge used for your artifacts if it is related to external parties and your team demands it. It can be algorithm descriptions or results of external projects. External means there is no cooperation or influence to the external parties. Implied by 3, no outside dependencies are needed, the other parts of the definition do not introduce additional dependencies.	3	Give the (kind of) knowledge and the representation of the knowledge needed for your team to work but is not (completely) present in your team. Is it needed to teach your team the usage of tools, is it required to transfer the domain knowledge to your team (by teaching), etc.? The description does not lists knowledge not yet part of the team that is needed.	1, 2, 3, 4, 5, 6	Give the roles in your team that handle dependencies. Give the role that has to learn knowledge or use an external artifact. No knowledge is given in this row, so no role is needed.		Give dependencies in executing the management process. Is the ongoing of your team managed by an external party (e.g., the product owner (SCRUM) is not part of your team)? Nothing to manage, so no process needed.		Give dependencies of process information in this row. For example, your team collects the information when external parties deliver knowledge or how much is billed by external parties. Not demanded by the definition.	6

Figure A5. Description of software development toy example based on the ßMACH protocol.
Presented is a set of key aspects. (See Fig. A3 and https://doi.org/10.5281/zenodo.10992007 for more details.)

	Product Knowledge		Demanded Knowledge		Roles		Process Knowledge		Process Information	
Inside Interfaces Give the management of the interfaces your team uses. It can be descriptions of already fixed interfaces, the (kind of) documents your team creates, or what you have to negotiate with cooperating teams (related to the internal border). Interfaces always mean interfaces of artifacts your team manages. Otherwise, it is a dependency.	Describe the realization of the interfaces (product representation). Describe what is known about the interfaces, which artifacts are related, how to give the technical specification, etc. Not mentioned in definition (the magical box makes everything).	2	Give the (kind of) knowledge not present in your team to manage the interfaces, e.g., if your team (or someone else) develops the interface definition in the management process. No interfaces, so no missing knowledge in the team.		Give the roles in your team that care about the interfaces. Give the role that has to learn knowledge or use an external document. No interfaces, so no roles needed.		Give the process to manage the interfaces given in this raw. Is it needed to negotiate, is it demanded to discover interfaces (missing documentation), how to realize the interfaces, how and when are interfaces defined or altered, how to perform testing, etc? No interfaces, so no process needed.		Give the information that has to be collected while performing the management process in this row. Is billing relevant information needed, is interface documentation (API description) recorded, is documenting different versions with time stamps mandated, etc? Not demanded by the definition.	6
Outside Interfaces Give the management of the interfaces your team uses with external parties (the real world). It can be descriptions of already fixed interfaces, the (kind of) documents you have to create later on, or what you have to negotiate with external parties (related to the external border). Examples are user interfaces, machine interfaces, and system interfaces. Interfaces mean always interfaces of artifacts your team has to manage. Otherwise, it is a dependency.	Describe the realization of the interfaces (product representation). Describe what is known about the interfaces, which artifacts are related, how to give the technical specification, etc. Not mentioned in definition (the magical box makes everything).	2	Give the (kind of) knowledge not present in your team to manage the interfaces, e.g., if the management process has to develop the interfaces. No interfaces, so no missing knowledge in the team.		Give the roles in your team that care about the interfaces in this row. No interfaces, so no roles needed.		Give the process to manage the interfaces in this row. Is it needed to negotiate, is it required to discover interfaces (missing documentation), how to realize the interfaces, how and when are interfaces defined or altered, etc? No interfaces, so no process needed.		Give the information to collect about the interfaces in this row. Is billing relevant information needed, is interface documentation (API description) demanded, is it mandated to document different versions with time stamps, etc.? Not demanded by the definition.	6
Inside Responsibilities What is your team related to cooperating teams responsible for? How is it possible to measure the success of your team? Which of the described requirements is related to your team? How are shared responsibilities managed, etc.?	Give the responsibilities related to the product, e.g., based on requirements. 3 successes that no responsebilities are given, the other parts of the definition do not introduce additional dependencies.	3	Give the responsibility for demanded knowledge, how to find out if your team has acquired it. The magical box (holding the knowledge) is shared with cooperating teams, so we have to return it.	4	Give the roles in your team that have to manage the measurement and achievement of the team's responsibilities to cooperating teams. To return the magical box does not demand a special role.	5	Give the process to manage the responsibilities in this row. Do you have regular meetings, plans in case of not achieving goals, etc.? Returning the magical box is a task that das not need management (the persons in the team know what to do).	5	Give the information to the responsibilities your team manages in this row. Is it needed to collect information about the process, is it mandated to notice the fulfillment of responsibilities, what kind of documents to use, etc.? Not demanded by the definition.	6
Outside Responsibilities What is your team related to external parties responsible for? How to measure the team's success (combined with the cooperating teams)? Which of the described requirements is related to your team? How are shared responsibilities managed, etc.?	Give the responsibilities related to the product, e.g., based on requirements. 3 successes that no outside responsibilities exist, the other parts of the definition do not introduce additional responsibilities.	3	Give the responsibility for demanded knowledge. How to find out if you have acquired it? The description does not lists knowledge not yet part of the team but is needed.	1, 2, 3, 4, 5, 6	Give the roles in your team that manage the measurement and achievement of the teams' responsibilities to external parties. No knowledge is given in this row, so no role is needed.		Give the process to manage the responsibilities in this row. Do you have regular meetings, plans in case of not achieving goals, etc.? Nothing to manage, so no process needed.		Give the information to the managed responsibilities in this row. Does your team need to collect information during the process? Is it mandated to notice the responsibility fulfillment? What kind of documents to use, etc.? Not demanded by the definition.	6
External Artifacts What kind of artifacts from an external party are used by your team? (Not external artifacts represented by product properties or dependencies.)	Give the external artifacts your artifacts demand. Are you using an external library, need an external service, is a component delivered by an external party, etc.? Give what your team needs to know about such artifacts. Not mentioned in definition, the magical box manages everything.	2	Give the knowledge about external artifacts not known by your team. Is an external party integrating an external artifact, does only an external party know what such an artifact does or how to use it, etc.? Not mentioned in definition, the magical box manages everything.	2	Give the roles in your team that handle the external artifacts. No external artifacts are given, so no role is needed.		Give the management process to handle external artifacts. No external artifacts are given, so no management process is needed.		Give the information to collect about external artifacts. Are you billed for using the artifacts, when are the artifacts delivered, when have you used an external service, how to collect this information, etc.? Not demanded by the definition.	6

Figure A6. Description of software development toy example based on the ßMACH protocol. Presented is a set of key aspects. (See Fig. A3 and https://doi.org/10.5281/zenodo.10992007 for more details.)

Affiliations

Marcus Hilbrich
Humboldt-Universität zu Berlin, Department of Computer Science, Berlin, Germany, marcus.hilbrich@informatik.hu-berlin.de

Ninon De Mecquenem
Humboldt-Universität zu Berlin, Department of Computer Science, Berlin, Germany, mecquenn@informatik.hu-berlin.de

Received: 5.05.2023
Revised: 10.04.2024
Accepted: 15.04.2024

Computer Science • 25(4) 2024 https://doi.org/10.7494/csci.2024.25.4.5997

Le Akanksha Bali
Kuljeet Singh
Vibhakar Mansotra

EYE DISEASE SEGMENTATION USING HYBRID NEURAL ENCODER DECODER BASED U-NET HYBRID INCEPTION

Abstract *Diabetic retinopathy (DR) is one of the major causes of vision problems worldwide. With proper treatment, early diagnosis of DR can prevent the progression of the disease. In this paper, we present a combinative method using U-Net with a modified Inception architecture for the diagnosis of both the diseases. The proposed method is based on deep neural architecture formalising encoder decoder modelling with convolutional architectures namely Inception and Residual Connection. The performance of the proposed model was validated on the IDRid 2019 contest dataset. Experiments demonstrate that the modified Inception deep feature extractor improves DR classification with a classification accuracy of 99.34% in IDRid across classes with comparison to Resnet. The paper Benchmark tests the dataset with proposed model of Hybrid Dense-ED-UHI: Encoder Decoder based U-Net Hybrid Inception model with 15 fold cross validation. The paper in details discusses the various metrics of the proposed model with various visualisation and multifield validations.*

Keywords fundus images, UNET, deep learning, diabetic retinopathy (DR), Indian Diabetic Retinopathy Image Dataset (IDRID)

Citation Computer Science 25(4) 2024: 579–620

1. Introduction

Blindness and vision loss can result from diseases of the fundus. A few frequent fundus conditions that impair vision are cataract, diabetes retinopathy, and age-related macular degeneration [4]. Diabetes is a serious pandemic worldwide, especially in Indian culture and around the world. As a result, there is an excess of diabetes, which spreads disorders like DR.

Diabetes has affected more than 422 million people worldwide out of which India ranks in the top 3 nations in terms of the number of diabetics. Half of the population, which has increased from 108 million to 422 million in recent years, now resides in India, China, the USA, Brazil, and Indonesia [11]. Diabetes can cause an eye condition called diabetic retinopathy. It takes place when high blood sugar levels harm the blood vessels in the retina, causing visual issues. Although diabetes can result in several problems, this paper will concentrate on diabetic retinopathy. After having diabetes mellitus for a while, diabetic retinopathy, an eye condition, develops. It harms the tiny veins in the retina, resulting in obstruction, damage, and erratic vein growth. The narrowing of the blood vessels in the eyes is a secondary effect of diabetes brought on by high blood sugar levels. A reduction in the blood flow to the retina may cause blindness. Among the symptoms of DR are lesions that develop on the retinal surface [7]. The vessels that carry blood in the retina are harmed by sustained high blood sugar levels, which results in diabetic retinopathy. This results in weaker arteries, obstructions, and fluid leaks. The body may react by developing abnormal new blood vessels, which can bleed and cause more harm. Its advancement is aided by factors such as elevated blood sugar, high levels of blood pressure and cholesterol as well as chronic inflammation. For early identification and treatment, periodic vision tests and managing diabetes are essential. The small blood vessels in the retina are harmed by high blood glucose levels, which results in DR. The macula enlarges and thickens because of the retina producing a greater amount of the blood, cholesterol, the latter of other lipids [28]. As sample blood is given to the retina, IrMAs (Intraretinal Microvascular Abnormalities) start to form as new, abnormally delicate blood vessels [12]. The rising pressure within the eye may potentially cause damage to the optic nerve. The most efficient method for identifying early indications of anomalies in the present clinical diagnosis is routine screening of diabetes individuals using fundus examinations. Blindness can be prevented if DR is identified in its earliest stages and treated right away [37].

The figure beneath shows the different stages of DR. As shown in the figure below, there are various stages of diabetic retinopathy. Fluid leakage and microaneurysms are features of mild non-proliferative retinopathy. Blockages in the blood vessels, edoema, and distortion are symptoms of moderate non-proliferative retinopathy. There are more blocked vessels because of severe non-proliferative retinopathy. A characteristic of proliferative retinopathy is the development of aberrant blood vessels, which can result in hemorrhage and retinal detachment. Due to the many fundus imaging tools, accurately grading and identifying diseases might be difficult. It may

also be challenging to distinguish between normal and pathological regions due to occlusion, shadow, reflection, or inadequate illumination in DR fundus pictures (Fig. 1).

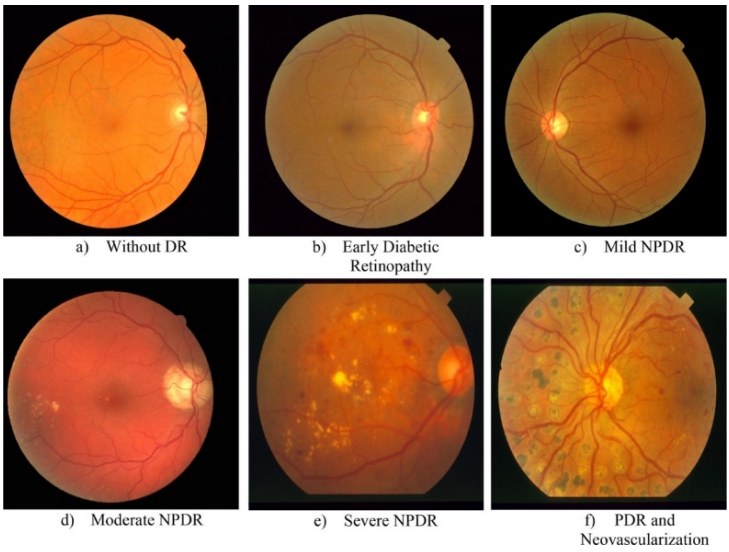


Figure 1. Dataset visualisation

Due to the many morphologies, sizes, and colours of lesions, manual retinal lesion assessment is laborious and demands great accuracy. Computer-assisted diagnostic (CAD) tools make it possible to diagnose diabetic retinopathy (DR) accurately and quickly, which helps with choosing the best course of action. To examine the Region of Interest (ROI) and determine the grade and severity of the disease, DR lesion segmentation is essential. The conditions that received the most research attention were glaucoma, AMD, and diabetic retinopathy [6]. Because ML techniques are more precise than traditional segmentation methods that rely on few indicators, they have taken their position. In comparison to conventional ML algorithms, DL, a contemporary technology, has demonstrated improved performance in DR lesion classification and segmentation [5].

In recent years, deep neural networks and particular machine learning models have shown promise in a variety of computer vision applications, particularly in medical picture analysis for diabetic retinopathy. Deep learning-based computer-aided diagnostic (CAD) systems that can categories anomalies can aid in medical decision-making and enhance patient care [25]. The rest of the paper is structured as follows. Related works on DR image classification are presented in Section 2. The information and suggested approach are explained in Section 3. The experimental analysis is presented in Section 4 and discussion are presented in Section 5. Section 6 presents the conclusions at the end.

2. Literature review

Depending on their study focus and area of interest, researchers in the subject of diabetic retinopathy have accomplished a lot of work. Researchers have suggested and applied a variety of machine learning techniques, as evidenced by the linked work in the fields of medical sciences and machine learning, however for diabetic retinopathy, a comparative analysis of these deep learning techniques is still absent. The work completed demonstrates that a fresh technique can be used when looking at the outcomes and comparing different machine learning algorithms for DR. In addition to several other analysis approaches, [22] used image processing for the automated and early detection of diabetic retinopathy. In their study, [40] suggested a method for improving images based on morphological operations, along with threshold-centered static wavelet transforms for the retinal fundus image and CLAHE (Contrast Limited Adaptive Histogram Equalization) for vessel enhancement. The method used by [50] is centered on a mixed classifier that detects retinal lesions through preprocessing, lesion extraction from candidates, feature formulation, and classification. The research leads to a further development of the m-Mediods-based modelling methodology, integrating it with the Gaussian Mixture Model to create a hybrid classifier that will increase classification accuracy. [2] investigated whether neural networks could recognize the signs of diabetes present in fundus photos and compared the network against a set of fundus images used for ophthalmologist screening. The research demonstrated the ability to find vessels, exudates, and hemorrhages. Their network achieved greater accuracy for the identification of diabetic retinopathy as compared to ophthalmologists. [15] contributed to the decrease in the number of features needed for the lesion categorization using feature ranking and Adaboost in their study. They suggested a novel two-step hierarchical classification method in which false positives or non-lesions are eliminated in the first stage.Bright lesions are further divided into two categories in the second stage: cotton wool patches and hard exudates. Additionally, red lesions are still categorized as haemorrhages and micro-aneurysms (MA).

Guo et al. [18] suggested a multi-scale feature fusion technique to address tiny lesion detection problems. To improve sensitivity, binary cross-entropy (BCE) loss alongside balancing coefficients was applied. The computational model was trained using full-resolution photos that were scaled to pixels without any additional preprocessing. Yan et al. [47] presented a mutually local-global U-Net to address the shortcomings of patchwise training, which fails to capture global context.

Other methods researchers have developed for segmentation purposes over the years like Fuzzy C-Means (FCM) clustering algorithms which are commonly used to divide image pixels into diverse cluster [30], and others like region growing methods are used to form distinct image regions based on some uniformity criteria such as grey level and colour [39], and mathematical morphology operations are performed by analysing the geometrical makeup of certain retina components [41,43]. However, most existing approaches are limited to acting on a single retinal image, making it impossible to extract spatial and spectral characteristics across many spectral slices at the same time.

Various researchers have advocated various segmentation strategies. In any event, these approaches only work on fundus pictures and have no pathological implications [14]. It is challenging to segment the vascular vessel treemap without discontinuities. Azzopardi et al. [3] created a filter called the Bar-selective Combination of Shifted Filter Responses (BCOSFIRE). Its parameters have an impact on the filter's performance. To detect retinal vessels, Wilfred Franklin and Edward Rajan [13] devised the Multilayer Perceptron Neural Network. For its representation, the weight of the feedforward network is altered using the backpropagation technique. Because it is dependent on pixel processing, it has a lower level of accuracy of 95.03%. Partovi et al. [31] proposed a model in which photos were classified using an error-based autonomous network. It was validated using a dataset of remote sensing photos. In the categorization of medical images, the Deep CNN (DCNN) model provided an extended feature extraction-classification technique. Gulshan et al. [17] trained the Deep Convolutional Neural Network (DCNN) to detect DR in retinal fundus images. Deep learning can be utilised in retinal fundus pictures to construct an algorithm that automatically detects DR and diabetic macular edoema. The specificity and sensitivity of the method for assessing DR expressed as moderate or worse DR or both were generated based on the key decision of the ophthalmology team. The method, which has 96.5% sensitivity and 92.4% specificity, was created using deep convolutional neural networks and a large quantity of data in various grades per image. By learning a deconvolution network, the authors of (Noh et al., 2015) [26] suggested an improved semantic segmentation technique. The convolutional layers are adopted from VGG16 as well, and the deconvolution network is made up of deconvolution and unpooling layers that identify pixel-wise class labels and forecast segmentation masks. On the same PASCAL VOC 2012 dataset, the proposed technique achieved 72.5% IoU. Shen et al. [38] introduced a multicrop pooling method that was used in DCNN to capture object salient characteristics in order to classify lung nodules on CT images.

In the literature [20], DR pictures were classified using a Gabor filtering and SVM classification model. Before employing the classifier, the Circular Hough Transform (CHT) and CLAHE models were fed with input images, yielding a detection rate of 91.4% on the STARE dataset. [36] made a contribution to the decrease in the amount of features needed for the lesion categorization using feature ranking and Adaboost in their study. They suggested a novel two-step hierarchical classification method in which false positives or non-lesions are eliminated in the first stage. Bright lesions are further separated into two categories in the second stage: cotton wool patches and hard exudates. Additionally, red lesions are still categorized as hemorrhages and micro-aneurysms (MA). For the diagnosis of various retinal abnormalities, the division of the blood vessel represents a crucial prerequisite. The accuracy of the cascaded U-Net architecture used by [36] on the DRIVE dataset was 96.92%, and its precision on the STARE dataset was 97.40%. [24] propose a deep CNN network that simultaneously slices arterioles and venous vessels from the DR pictures. The DRIVE dataset, on which the framework was tested, achieved 87% sensitivity and 98% specificity.

Data augmentation using the U-Net model was done over the DRIVE dataset by [46] which obtained an AUROC score of 0.97 for segmenting blood vessels.

On the MESSIDOR dataset [37] classified DR fundus images using a modified AlexNet architecture by applying the appropriate pooling, softmax, and ReLU layers. The MESSIDOR dataset showed 96.6% accuracy for the suggested model. Xiancheng et al. [45] demonstrated a DR model for classification using the InceptionV3 as architecture and a short dataset transfer learning approach. Utilizing an SGD optimizer including the cosine loss function for binary classification, Hagos' technique obtained 90.9% accuracy. In a study by [19], the referable lesion regions in DR images were examined using a generalization of the backpropagation approach and a poorly supervised model. On the Kaggle dataset and the E-Ophtha dataset, [33]'s suggested method has an area around the ROC curve (AUC) of 95.50% and 94.90%, respectively. An ensemble of models trained with deep learning was used in a study by [29] to detect red lesions in fundus images. In that technique, 3232 pixel patches were extracted initially and put into a Deep CNN. The random forest (RF) classifier also received hand-crafted features that were extracted in addition. Orlando's approach demonstrated how a hybrid feature vector with both hand-crafted and deep learning-based features might enhance the networks' performance and obtain 89.32% AUC. In order to classify eye illnesses, Bali et al. [8] suggested a DFex-hybrid strategy combining the BeeHive model, CGAN, and PSO. In the RFMiD and ODIR datasets, they obtained, respectively, accuracy of 98.79%, sensitivity of 95.99%, specificity of 99.79%, and accuracy of 97.16%, F1 score of 96.81%.

The goal of this paper is to develop an automatic diagnosis method for DR. The paper discusses multiple deep learning techniques for the segmentation of the fundus image as shown in Table 1. The paper is benchmarked over main dataset of IDRID for respective classifications. The paper is capable of not only reducing the computation time for the enhanced Dense-ED-UHI: Encoder Decoder based UNet Hybrid Inception architectures but also is able to hybridize it with inception V3 to calculate complex features for better extractions.

Considering the aforementioned review, we developed a unique architecture to overcome the aforementioned restrictions. The following is a description of the main contributions of the suggested research efforts:

1) This study uses a customised U-Net model (UHI) to semantically segment retinal lesions. The encoder used in the model is a deep network encoder called "InceptionV3". The smaller convolutions performed by the InceptionV3 model allow for a quicker training procedure.
2) To resize the convolution nearest neighbour, the model applies a unique upsampling technique that involves pixel-wise periodic shuffling convolution. Compared to traditional approaches, our procedure speeds up network convergence and creates retinal images with excellent resolution that are free of checkerboard artefacts.

3) Comparing the suggested model to other existing works, it produced state-of-the-art results for the diagnosis of DR lesions.
4) A hybrid U-Net architecture with inception design and multiple kernel extractions is used in the study to infect deep encoding.
5) Finally, the study evaluates the model's performance across multiple classes.

Table 1
Literature review comparison

Study	Methodology	Advantages	Disadvantages	Specificity	Recall	AUC
[17]	DL: Convolutional Neural Network	High accuracy and precision	Large labelled datasets required	0.99	0.90	NA
[1]	DL: Ensemble of CNNs	Automated feature learning	Computationally intensive	0.87	0.96	0.98
[16]	DL: CNN	Handles diverse lesion morphologies	Prone to overfitting	0.98	0.94	0.97
[42]	DL: CNN	Robust to noise and variations	Data augmentation may be required	0.91	0.90	0.93
[49]	ML: Random Forest	Can handle small or imbalanced datasets	Parameter tuning may be required	0.70	0.80	NA
[34]	ADL (Active Deep Learning)	Can be used to define the level of severity of retinal images important patches with provision of (ROI).	Can be improved through data mining and data exchange methods	0.95	0.92	NA
[48]	Deep DR	It is useful for DR severity classification	The model is too complex	0.97	0.97	NA
[35]	Deep learning methods	It assigns pixel values for determining the importance and evaluation of forecasted category	Parameter tuning may be required	0.90	0.90	NA
Proposed Model	UHI Network	Described in section 3	After Hyperparameter Tunning	0.997	0.989	0.999

3. Methodology

The methodology, experiment design, and workflow employed in the proposed framework in Figure 2 were used in this section's full explanation of the dataset. The following shown below is the proposed architecture

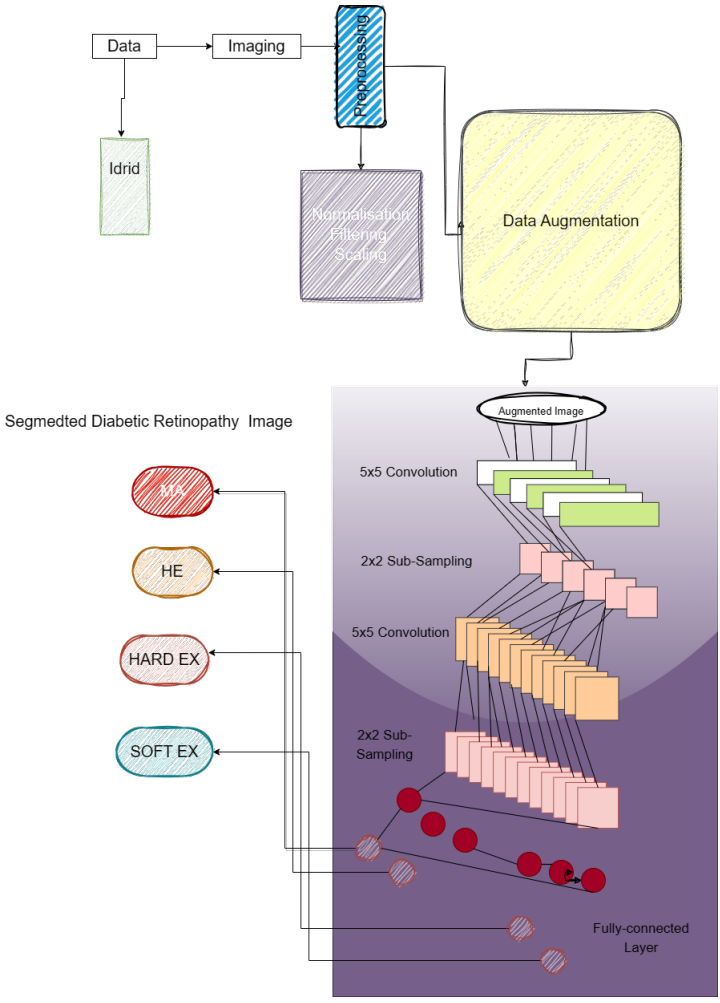


Figure 2. Flow diagram

3.1. Dataset rescription

The open-access Indian Diabetic Retinopathy Image Dataset (IDRid) focuses on diabetic retinopathy, an eye-related consequence of diabetes. The dataset is intended to support study and advancement in the area of retinal analysis of images, particularly for the detection and grading of diabetic retinopathy. There are 516 retinal pictures

in all, each with a resolution of 4288 × 2848. The pixel-level annotation has provided a binary mask for each distinct DR deformity, such as haemorrhages, hard-exudates, soft-exudates, and microaneurysms. Furthermore, every 516 photos have been graded and given the DR severity [21]. The data description in Table 2.

Table 2
Data description

S. No	Keypoints	Description
1	Content	The IDRid dataset includes fundus photographs of the retina. 516 high-resolution retinal pictures make up the entire collection
2	Annotation	Expert-verified ground truth labels describing the existence and severity of diabetic retinopathy are added to each image in the collection as annotations. The labels, which vary from 0 to 4, represent the International Clinical Diabetic Retinopathy severity levels from "No DR" (zero) to "Proliferative DR" (four)
3	Image formats	The uncompressed TIFF format used by the IDRid dataset preserves the high-resolution details required for analysis and diagnosis
4	Dataset division	A training set as well as a testing set were created from the dataset. There are 413 images in the training set and 103 images in the testing set. This category enables researchers to assess how well their algorithms function on hypothetical data
5	Terms of usage	In accordance with the rules of the Creative Commons Attribution 4.0 International (CC BY 4.0) licence, the IDRid dataset is available for research purposes. This means that, as long as the original authors are properly credited, you are free to use the dataset, modify it, and redistribute it for any non-commercial reasons

3.2. Implementation details

For model training, a GPU with 8GB of RAM is used. In the papers of [9, 10], both attempted were some of the earliest papers to utilise U-Net for image segmentation. In these studies, it is discussed how to segment retinal lesions in images of diabetic retinopathy using the U-Net design. Since then, the U-Net architecture has been widely used for a variety of medical picture segmentation applications, including the identification of diabetic retinopathy. After being initially presented for broad biomedical image segmentation tasks, the U-Net architecture has now been adopted and used in a number of medical imaging areas, including diabetic retinopathy.

While segmenting retinal structures and lesions in images of diabetic retinopathy has shown the U-Net architecture to be useful, there are certain restrictions as well. Utilising U-Net for diabetic retinopathy has some drawbacks, including a restricted

capacity to capture global context. U-Net may not successfully consider greater spatial connections and contextual signals because of its encoder-decoder architecture, which places a heavy emphasis on local information. This constraint may reduce the model's capacity to incorporate global data necessary for precise categorization of diabetic retinopathy, which may have an adverse effect on the model's general efficacy. Additional architectural alterations or the incorporation of other methodologies might be required to improve the integration of global context and increase the categorization accuracy in order to counteract this limitation.

The study adopts concepts from the Inception design and inserts them into the neural network instead of using the conventional U-Net architecture. The detection becomes simpler as Inception design combines numerous kernel sizes to collect information on various scales and at various granular levels. Additionally, U-Net Inception improved feature representation, led to a more thorough knowledge of the retinal image, and may improve the performance of the suggested model as a whole. Additionally, the expansive to contract cross-linking helps produce the extremely precise segmented output image. There are four blocks of encoder units in the U-Net contract chain. Each encoder unit has two convolution layers that are followed by a max-pooling layer. Every time a pooling process is used, the feature elements are doubled. The bottleneck, which consists of two convolution layers with one at the top, is the essential component that separates contractile and expanded techniques. A U-Net expanding path is being performed by four decoder units and consists of two convolution layers, a de-convolution layer, and two comparable feature maps using the contract trail. The Table 3 below discusses the brief difference between the both.

The proposed work with the dataset split into the training, validation, and testing set for DR classification in retinal pictures. In the Figure 3 seen below, the paper proposes a detailed overflow of utilising four benchmark datasets to train the images, follow the methods and update if required. Following this, testing the dataset is done and if the model performs up to the expectations, it can be utilised in the field of medical imaging. The Figure 3 and Figure 4 describe the primary steps of data preprocessing the the following order

1) Training – Testing and Validation Split, as discussed, the training data 70% is used to train and reiterate the model. Testing and validation process used to analyse the fitting parameter and ensure precise modelling.
2) Data preprocessing, it includes various steps like Normalisation of the given input sequences, Filterining Scaling etc.
3) Mask Encoding or One Hot encoding is done to create a binary mask for multiple classes used.
4) Model creation, hyperparameters and other units are discussed below.
5) Post training cycle is completed the model goes through a vivid evaluation for Cross validation across all metrics such as accuracy, precision, recall, sensitivity, specificity.

Table 3
Difference between U-Net Inception and Standard U-Net

Aspect	UHI: U-Net Inception	Standard U-Net
Architecture	The U-Net framework incorporates concepts from the Inception architecture	Encoder-decoder as well as skip connections are part of the original U-Net architecture
Multi-scale feature extraction	Uses various kernel sizes to efficiently capture data at various scales	Focuses heavily on skip connections for the extraction of local features
Global context	Due to the incorporation of modules inspired by Inception, the global context may be captured more successfully	Capability to capture the global context is somewhat limited
Fine detail representation	Possibility of enhanced fine detail representation in retinal pictures	May rely more heavily on skip connections to express fine details
Performance potential	Possibility of improving categorization accuracy by using data from multiple scales	Established performance on tasks involving the classification of diabetic retinopathy
Complexity	May add to the complexity of the architecture and processing	Architecture that is less complex and contains fewer extra parts

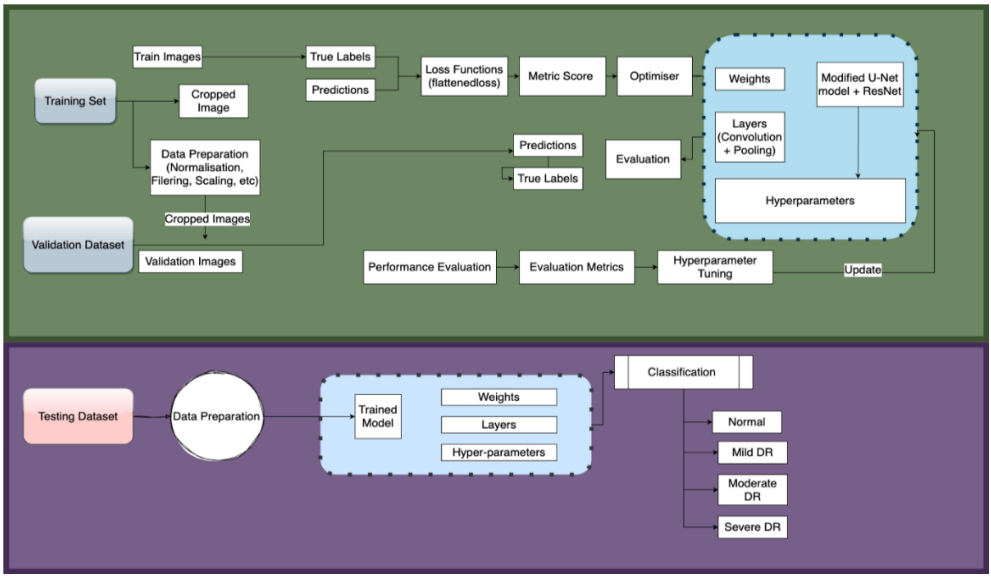


Figure 3. Data split and implementation

The figure below talks about the proposed U-Net Inception model for diabetic retinopathy classification. The primary aims the figure postulates to demonstrate is that training and testing sets are passed via various preprocessing pipelines after which the train images are segregated with loss function and back propagation utilities of the deep network inclusive of weights, biases, layers and hyperparameters as in Figure 4.

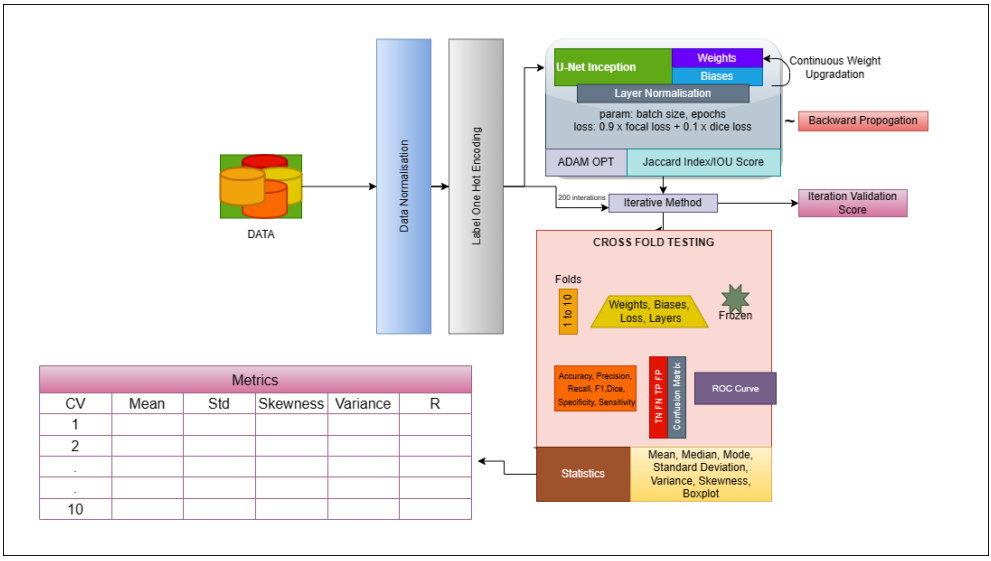


Figure 4. Data evaluation and cross validation set

The following is a description of the various building blocks' intended architectural details.

3.2.1. Encoder unit

The encoder is essential for capturing and extracting hierarchical characteristics from the input retinal pictures in the U-Net Inception Net structure for diabetic retinopathy classification. It is made up of convolutional layers which combine input data with filters to extract features. The encoding layers steadily increase the filter depth, enabling the neural network to capture intricate information important for categorising diabetic retinopathy.

The encoder uses down sampling techniques like max pooling or stride convolution to increase the number of feature channels while decreasing the spatial dimensions of the feature map. This down sampling aids in obtaining more advanced depictions of the input and summarizing the data. The incorporation of the Inception module into the encoder is one of the unique features of the U-Net Inception Net. The Inception module includes of a max pooling branch and parallel convolutional branches that use different kernel sizes, notably 1×1, 3×3, and 5×5. These branches catch

features at many scales and assist the network in simultaneously learning local and global knowledge. The encoder of the U-Net Inception Net improves the network's capacity to collect complex and multi-scale information that are necessary for precise diabetic retinopathy categorization by including the Inception module. Overall, the U-Net Inception Net encoder collects pertinent and discriminative characteristics from the input retinal images, creating a rich representation which is further used in the later phases of the architecture for classifying diabetic retinopathy.

3.2.2. Inception unit

The U-Net Inception Net design for classifying diabetic retinopathy includes the Inception module, which was first introduced in GoogLeNet. By obtaining multi-scale and multi-level features, it improves a network's feature extraction capabilities. The Inception module includes of a max pooling branch and parallel convolutional branches with various kernel sizes, such as 1×1, 3×3, and 5×5. The network may collect data at multiple scales since every branch captures characteristics at a distinct receptive field size. The 1×1 convolutions are in charge of dimensionality reduction and input channel reduction, which helps to reduce computational complexity.

The U-Net Inception Net is capable of recording local as well as global information by integrating the Inception module. The concurrent branches within the module allow the network to simultaneously learn complicated and varied information. This is especially helpful in the classification of diabetic retinopathy, because the network must collect information at numerous scales due to the presence of various lesions, including microaneurysms and exudates. The U-Net Inception Net's capacity to extract discriminative characteristics pertinent to diabetic retinopathy is improved by the Inception module. The network may collect both fine-grained minutiae and high-level semantic data by combining multi-scale features. This aids in increasing the classification task's accuracy and resilience, allowing for more precise detection and assessment of diabetic retinopathy (DR) in retinal pictures.

3.2.3. Skip connections

A key element of the U-Net design, which includes the U-Net Inception Net, for the classification of diabetic retinopathy is skip connections. These connections are essential for enhancing information transfer and maintaining spatial details across the network. Skip links create immediate connections among the encoding and decoding pathways in the U-Net Inception Net. These links make it possible to combine multi-scale along with multi-level features, which makes it easier to integrate low-level and high-level data. The feature maps from the respective encoder layers to the decoder layers are concatenated to create the skip connections.

The U-Net Inception Net successfully blends semantic detail from the decoder with low-level fine-grained features from the encoder by including skip connections. This gives the network a thorough comprehension of the linked features in the input retinal images. In addition to improving the network's capacity to precisely

classify diabetic retinopathy lesions, it aids in the preservation of spatial information. Additionally, skip connections help to solve the issue of a data bottleneck that might develop in deep networks. A network is able to access feature maps from various scales and levels by simply linking the encoding and decoding paths, which improves gradient flow and solves the vanishing gradient issue. The skip connections in the U-Net architecture Inception Net architecture, in general, promote efficient knowledge propagation, enable the integration of multi-scale data, and contribute to the precise identification of diabetic retinopathy lesions by fusing local information with global context.

3.2.4. Bottleneck

The bottleneck, which sits in the middle of the encoder and decoder paths, acts as a bridge to collect the input data in its most compressed and abstract form. The bottleneck in the U-Net Inception Net is often made up of several convolutional layers including Inception modules.These layers and modules are created to capture highly discriminative properties, which are essential for correctly classifying lesions associated with diabetic retinopathy.

The U-Net Inception Net gains the capacity to collect characteristics at various scales and levels of abstraction by including Inception modules within the bottleneck. The Inception modules' parallel convolutional branches make it easier to extract detailed information, allowing the network to pick up on intricate patterns and structures found in retinal images. The bottleneck's job is to compile the encoder's learnt representations and get them ready for the decoding path. It seeks to decrease the computational difficulty of the network while capturing the fundamental properties that are most pertinent to the classification task.

The bottleneck efficiently compresses the data, making the retinal pictures' representation more condensed and expressive. It allows the network to generate precise predictions throughout the succeeding decoding and classifying stages by collecting the most discriminative properties.

3.2.5. Activation function

In order to introduce non-linearity and improve the network's capacity to learn intricate correlations between the input data and the target labels, activation functions are crucial. The Rectified Linear Unit (ReLU), Leaky ReLU, and Exponential Linear Unit (ELU) are frequently used activation functions in the U-Net Inception Net. ReLU is a widely used activation function that maintains positive values while setting negative values to zero. Leaky ReLU avoids the dying ReLU problem by permitting a modest non-zero slope for negative values. ELU has a smooth curve for negative values, unlike ReLU, which does not. The non-linear mapping among labels for diabetic retinopathy (DR) and the input retinal pictures are simulated with the help of these activation functions. The U-Net Inception Net can capture complex patterns and data by incorporating non-linearity, which enhances its ability to distinguish between different types of diabetic retinopathy lesions and properly categorize them.

The selection of the activation function is influenced by the specific requirements for the classification task as well as the characteristics of the dataset.

3.2.6. Decoder unit

Upsampling the feature maps and regaining the lost spatial resolution are under the purview of the Decoder. The decoder for the U-Net Inception Net commonly consists of upsampling layers like bilinear interpolation or transposed convolutions. These layers expand the feature maps' spatial dimensions, enabling the network to recreate the retinal pictures' more minute details.

The decoder also includes skip connections that combine the feature maps from the associated encoder layers. In order to accurately classify data, these skip connections are essential for maintaining spatial data as well as combining low-level and high-level properties. The U-Net Inception Net's skip connections make it easier to combine multi-scale features, giving the network access to both fine-grained specifics and high-level semantic data. This helps gather pertinent data from many levels of abstraction and enhances the network's comprehension of diabetic retinopathy lesions. The decoder in the U-Net Inception Net successfully reconstructs the spatial details of the retinal pictures using upsampling operations and skip connections, providing the data required for precise diabetic retinopathy categorization.

3.2.7. Hyperparameter set

The performance of the U-Net Inception Net architecture for the categorization of diabetic retinopathy is strongly influenced by a number of hyperparameters. During optimisation, the learning rate defines the step size, which affects convergence speed and stability. The batch size regulates how many samples are handled during each training iteration, which has an impact on both the accuracy and efficiency of gradient estimation. The U-Net Inception Net's ability to learn complicated characteristics depends on how many layers it has. Deeper networks can capture more detailed patterns but need more computer power. Regularisation factors that avoid overfitting and regulate the amount of regularisation used include dropout and L1 or L2 regularisation. The model's non-linear behaviour showcased in Figure 5 and capacity to represent complicated relationships are significantly influenced by the selection of activation functions, such as ReLU, Leaky ReLU, or ELU.

Choosing the best hyperparameter values for diabetic retinopathy classification employing the U-Net Inception Net architecture typically requires experimentation and validation on a different validation set. Techniques like grid search or random search may be used to explore various configurations. The metrics used to evaluate the performance are iou_score, focal loss, dice loss, accuracy, binary accuracy, AUC and ROC, Specificity, Sensitivity.

a) b)

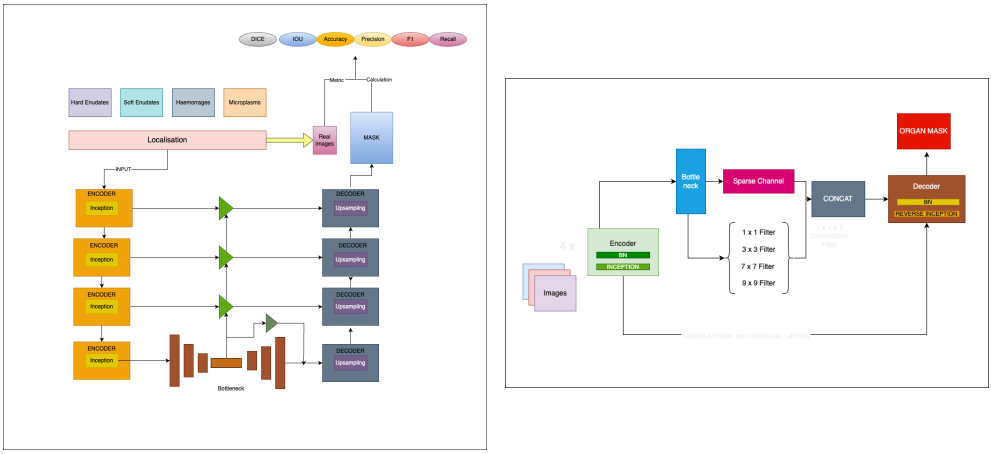


Figure 5. Data evaluation and cross validation set: a) Unet architecture; b) Inception V3 convolutional network

4. Results

4.1. IDRID

Specific kinds of diseases that are frequently seen in visual pictures include hard exudates, soft exudates, haemorrhages, and microaneurysms. These kinds of lesions are frequently a sign of several visual illnesses, such retinopathy caused by diabetes. An extensively used collection containing retinal pictures annotated for these abnormalities is called IDrid (Indian Diabetic Retinopathy Image collection). Let's talk about the traits associated with these tumours and the criteria needed for segmenting them.

4.1.1. Hard exudates

In the retina, triglycerides seep through injured veins, forming brownish or white plaques known to be hard discharges. They frequently take the form of tiny, rounded, or longitudinal tumours with distinct margins. As their buildup can cause blurred vision and act as a sign of the seriousness of diabetes-related retinopathy, these hard discharges must be segmented in order to be quantified and their development tracked.

4.1.2. Soft exudates

Puffy white or yellowish blemishes referred to as cotton fibre areas, soft discharges are the result of strokes of the ocular sensory fibre layer brought on by poor blood flow. These resemble amorphous tumours with hazy edges. Soft exudates must be segmented in order to determine their amount and location because they can be an indication of ocular hypoxia.

4.1.3. Haemorrhages

Blood leaks into the retinal cells during haemorrhages, which happen as veins burst. They resemble irregularly shaped, black lesions that are red or dark in colour. As they can reflect the extent of retinal harm and the evolution of diabetes-related retinopathy, haemorrhages must be segmented in order to determine their existence and pinpoint where they are.

4.1.4. Microaneurysms

These tiny, rounded expansions of the capillaries in the retina are believed to be the first symptoms of retinopathy caused by diabetes. They appear as tiny, rounded lesions that are brilliant red. Since the existence and growth of microaneurysms can suggest the possibility of progressing to more severe retinal stages, segmenting them is crucial for their measurement and management.

4.2. Retinopathy grade classification

From Table 4 and Figure 6 with a specificity of 0.909091 for type 0 retinal degeneration, cases who lack Retinopathy that can be identified with excellent precision. The F1-score value for third-degree retinopathy grade is 0.917197, which indicates that it is possible to forecast grade 3 retinopathy with an adequate level of recall as well as accuracy. Recall and sensitivity are used equally in this list. For example, the recall/sensitivity for retinopathy grade 2 is 0.972222, meaning that roughly 97.2% of cases of grade 2 retinopathy were effectively detected by the model. The detection rate for grade 1 retinopathy is 0.941176, meaning that the model correctly identified cases of grade 1 retinal with a 94.1% accuracy. Prediction of category wise classes is presented in Figure 7.

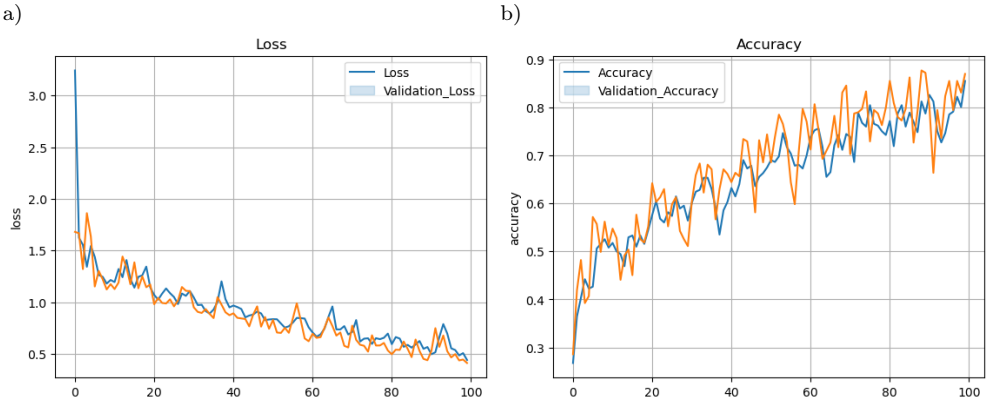


Figure 6. Loss and accuracy plot for Retinopathy Grade Classification: a) loss training validation plot; b) accuracy training validation plot

Table 4
Classwise metric distribution (testing)

Class	Accuracy	Precision	Recall or Sensitivity	F1-Score	Specificity
Retinopathy grade = 0	0.912621	0.955224	0.914286	0.934307	0.909091
Retinopathy grade = 1	0.941748	0.941176	1.000000	0.969697	0.842857
Retinopathy grade = 2	0.902913	0.897436	0.972222	0.933333	0.941935
Retinopathy grade = 3	0.873786	0.972973	0.867470	0.917197	0.900000
Retinopathy grade = 4	1.000000	1.000000	1.000000	1.000000	1.000000

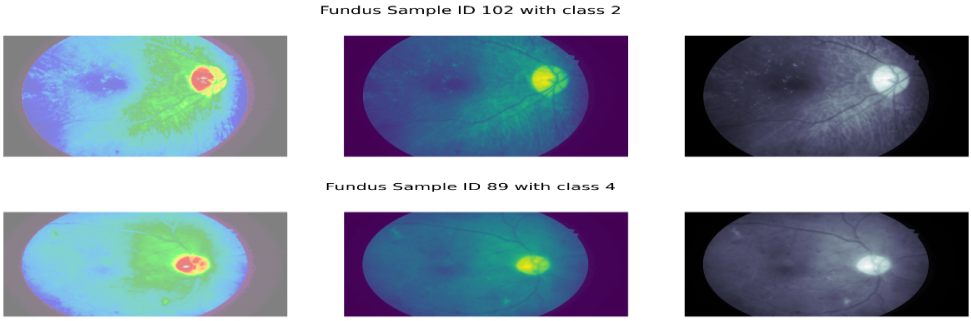


Figure 7. Class wise prediction

4.3. Classification risk of macular edema

The percentage of correctly predicted cases to all occurrences is what is meant by the term "accuracy", which measures overall competence. We can see the accuracy for every category in the following table. For example, the prediction accuracy for category 0 is 0.953995 (as per the training loss 8), meaning that in almost 95.4% of cases, the algorithm correctly predicted the likelihood of macular edema level 0. The model's prediction at classifying True Positives amongst the Predicted Classes that are positive is measured in the "Precision" row. For example, the precision for class 1 is 0.991736, indicating that the algorithm successfully identified cases of macula edema level 1 with a high precision of 99.2% (cf. Fig. 8).

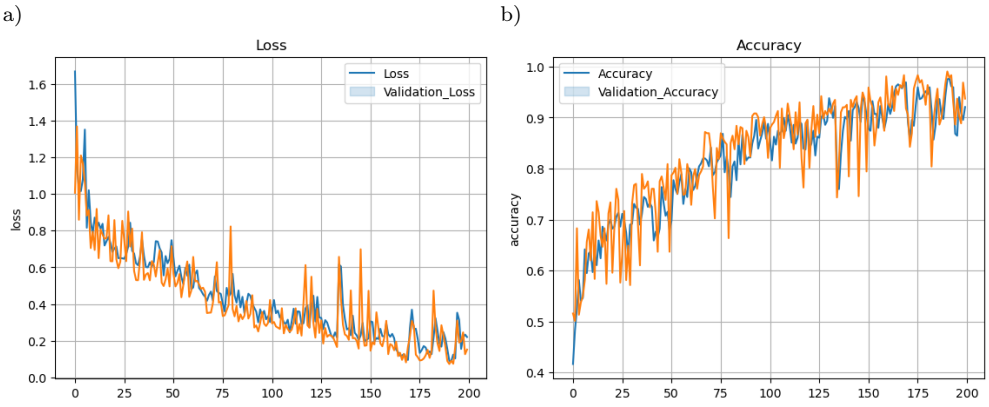


Figure 8. Loss and accuracy risk of macular edema: a) loss training validation plot; b) accuracy training validation plot

It is the same as sensitivities in Table 5 and 6. For example, the recall/sensitivity for category 2 is 1.000000, meaning that every incidence of macular edema level 2 was successfully detected by the algorithm, resulting in an ideal recall. Better performance is indicated by a higher F1 score. For instance, the F1 rating for class 1 is 0.979592, Indicating a strong equilibrium between recall and precision in classifying macular edema level 1. For instance, the specificity for class 0 is 0.971910, showing an elevated degree of reliability in detecting cases without macular edema at level 0.

Table 5
Class wise metric distribution (training)

Class	Accuracy	Precision	Recall or Sensitivity	F1-Score	Specificity
Retinopathy grade = 0	0.951456	0.950000	0.966102	0.957983	0.931818
Retinopathy grade = 1	0.961165	0.989011	0.967742	0.978261	0.900000
Retinopathy grade = 2	0.990291	0.981818	1.000000	0.990826	0.979592

Table 6
Class wise metric distribution (testing)

Class	Accuracy	Precision	Recall or Sensitivity	F1-Score	Specificity
Retinopathy grade = 0	0.953995	0.977876	0.940426	0.958785	0.971910
Retinopathy grade = 1	0.963680	0.991736	0.967742	0.979592	0.926829
Retinopathy grade = 2	0.956416	0.924051	1.000000	0.960526	0.907216

The class wise prediction is showcased in Figure 9.

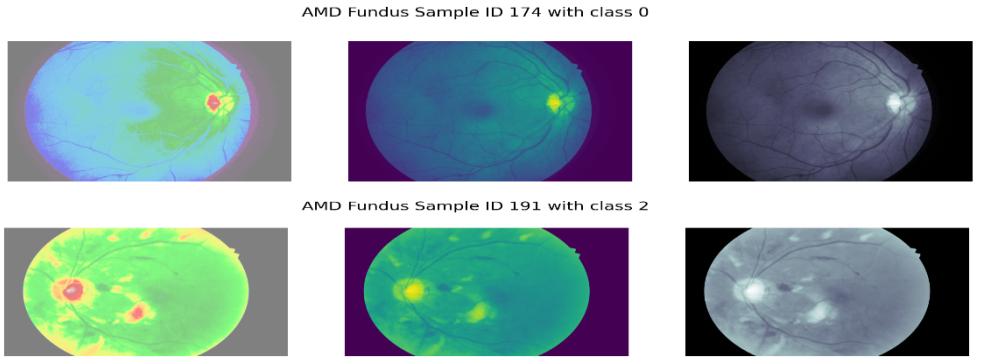


Figure 9. Edema class wise prediction

Table 7 highlights the distribution of samples of data throughout distinct categories of lesion types, possibly linked to medical imaging analysis. The table provides the total number of the samples for each given lesion type. "Hard Exudates" form the most prevalent class with 350 samples, next to "Soft Exudates" with 150 samples, "Hemorrhages" with 120 samples, and "Microaneurysms" with 110 samples. These values represent how often each lesion type occurs in the dataset, offering vital information into the dataset's makeup. Such information is critical for creating reliable artificial intelligence models, since the class distribution might affect the model's accuracy and generalization.

Table 7
Data samples per class

Lesion type	Number of samples
Hard exudates	350
Soft exudates	150
Hemorrhages	120
Microaneurysms	110

The provided table highlights the dataset's variety and identifies probable patterns or abnormalities in lesion incidence. In medical settings, this knowledge supports both scientists and clinicians in knowing the cause of certain problems, guiding diagnosis, treatment techniques, and future study. It is vital to remember that the application of these figures should be determined considering the unique clinical environment and the aims of the study, whether it's for boosting diagnostic accuracy or identifying disease features.

4.3.1. Hard exudates

The difference in the predicted and actual classes of hard exudates is measured using a loss operation, and a median loss of 0.057894 is the mean of the function throughout train. Better model scores in accurately representing the characteristics of hard exudates is shown by a smaller loss. The accuracy of the model appears to be stable throughout various training iterations based on Figure 10, Figure 11 and Figure 12, as indicated by a small variance of 0.014040. The amount of variance in the performance of the model is indicated by the least loss of 0.035471 to the largest loss of 0.090734. The model performs well overall in categorising pixels into either hard exudates or non-hard exudates, according to the mean accuracy of 0.999293. This shows that the model properly picks up on the characteristics that set hard exudates apart and can correctly distinguish these from various ocular components. The model's accuracy remains constant across assessments, as seen by the low standard deviation of 0.000162. The range of variation in the model's accuracy across several samples is highlighted by the least accurate value of 0.998942 and the highest accuracy of 0.999645.

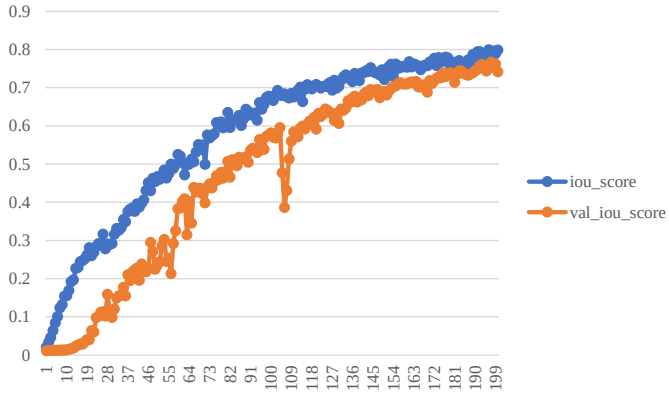


Figure 10. Training IOU_score

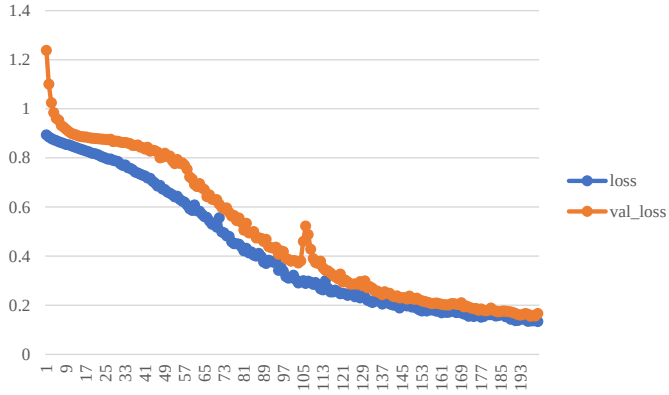
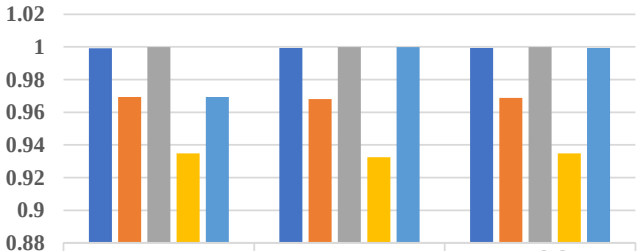


Figure 11. Focal+dice loss



	Train	Test	Valid
Accuracy	0.999257	0.999327	0.999326
Binary Accuracy	0.969272	0.968051	0.968795
AUC	0.999835	0.999873	0.999871
Specificity	0.934805	0.932559	0.934836
Sensitivity	0.969272	0.999873	0.999326

Figure 12. UHI model performance

Figure 13 and Figure 14 illustrates the prediction of Unet Inception and Resnet Models.

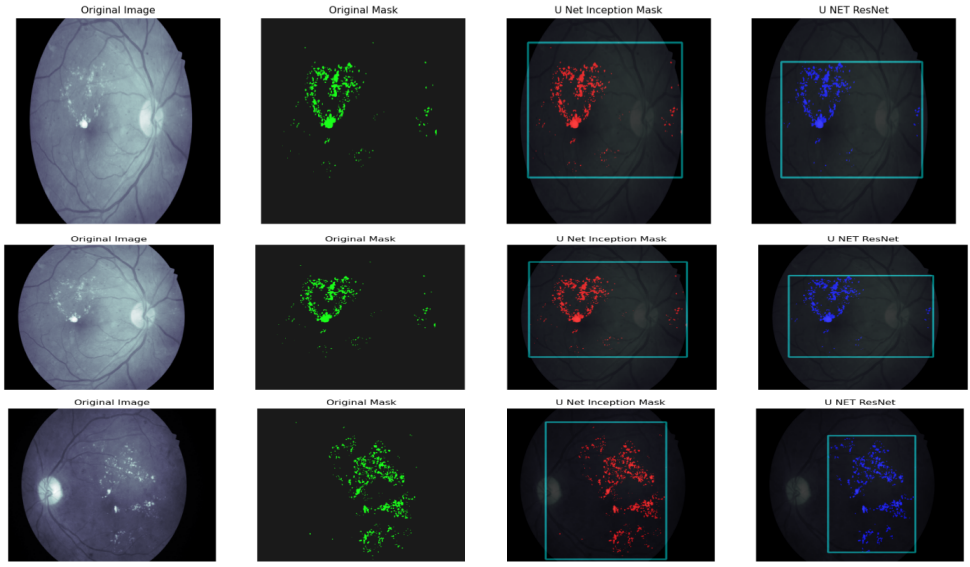


Figure 13. Mask prediction and BBox using Dense-ED-UHI:
Encoder Decoder based Unet Hybrid Inception (proposed model)

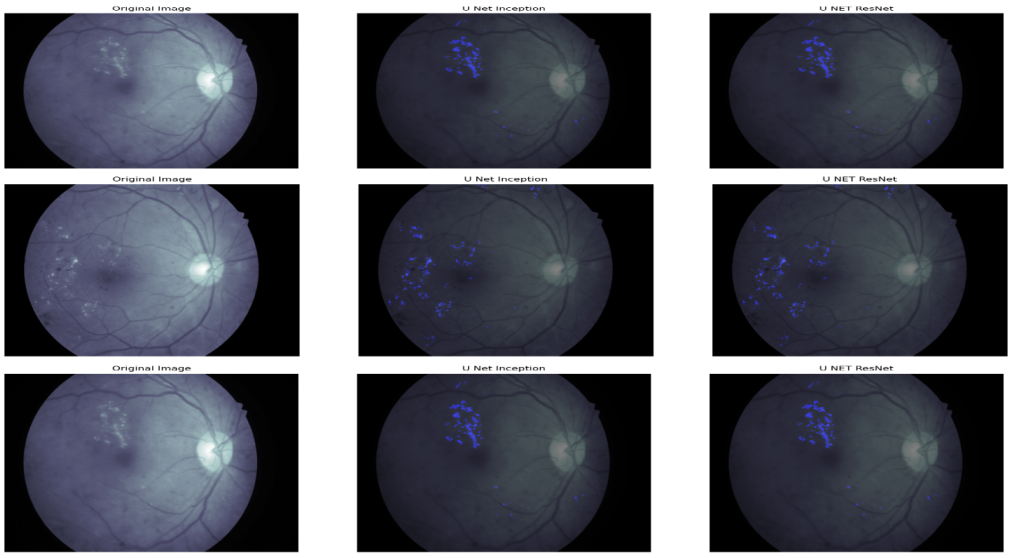


Figure 14. Comparison with other state of the art backbone (Resent50)

As per Table 8 the median amount of correctly anticipated hard exudates is represented by the median true positive value, which is 34,116.49. This score sheds light on the algorithm's precision in hard exudate detection and segmentation. The variance in true positives among various evaluations is shown by the standard variation of 13,782.728251. The range of variance in the model's ability to recognise hard exudates is shown by the smallest value of 11,497 and the highest number of 71,688.

Table 8
Statistical analysis

Statistical test	Loss	Accuracy	True positives	AUC	Specificty	Sensitivity
mean	0.057894	0.999293	34,116.490000	0.966475	0.999851	0.929403
std	0.014040	0.000162	13,782.728251	0.009651	0.000046	0.019548
min	0.035471	0.998942	11,497.000000	0.942636	0.999746	0.881022
25%	0.046062	0.999188	23,017.000000	0.959663	0.999817	0.915989
50%	0.057263	0.999304	31,707.500000	0.967367	0.999855	0.930924
75%	0.069872	0.999411	43,709.000000	0.974321	0.999885	0.945510
max	0.090734	0.999645	71,688.000000	0.982697	0.999944	0.962151

AUC (Area Under the Curve) whose value is 0.966475 represents the algorithm's overall efficacy in identifying hard and non-hard exudates. Stronger discriminating skills are indicated by a larger AUC, with numbers near 1 signifying outstanding

results. The model's capacity to accurately recognise non-hard secretions is demonstrated by its average specific of 0.999851. The algorithm's ability for correctly recognising hard exudate regions is indicated by a median sensitivity value of 0.929403. For assessing the model's effectiveness in accurately identifying hard exudates, the two metrics are crucial.

Overall, the statistical assessment of rigid exudates offers insightful information about the effectiveness of the categorization approach. The model is efficient in capturing the features of hard exudates, as evidenced by its minimal loss and excellent accuracy. The simulation is able to differentiate among hard and non-hard exudates, as evidenced by its elevated AUC and accuracy values. Yet, the model could at times overlook some hard exudates given the significantly reduced sensitivities. These data points act as crucial indicators for assessing and enhancing the categorization algorithm's efficacy for hard exudates in ocular pictures.

4.3.2. Haemorhages

Valuable information about its efficacy in identifying and categorising haemorrhages in retinal pictures may be learned from the statistical examination of the haemorrhages dataset. The model properly classifies pixels as haemorrhages or non-hemorrhages with an average accuracy and absolute accuracy of 0.999264, respectively. This shows the the simulation accurately depicts haemorrhages' distinctive characteristics and can separate these from various other ocular formations. The durability of the model is further supported by the low standard deviation of 0.000230 (as per Table 9), which shows that the correctness of the model is constant across several evaluations. The range of variance in the model's accuracy across several samples is highlighted by the least precision of 0.998607 and the highest confidence of 0.999657. Figure 15, Figure 16 and Figure 17 describe the training and UHI network on Haemorrhages dataset.

Table 9
Statistical analysis

Statistical test	Accuracy	Binary accuracy	AUC	Specificty	Sensitivity
mean	0.999264	0.999264	0.974184	0.999812	0.939016
std	0.000230	0.000230	0.008255	0.000072	0.018514
min	0.998607	0.998607	0.956402	0.999574	0.901336
25%	0.999184	0.999184	0.968939	0.999759	0.927588
50%	0.999331	0.999331	0.973982	0.999836	0.938238
75%	0.999403	0.999403	0.979575	0.999862	0.948270
max	0.999657	0.999657	0.990752	0.999934	0.976834

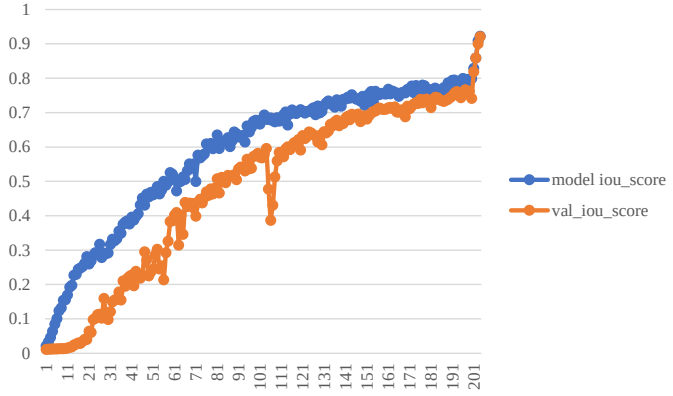


Figure 15. Training IOU_score

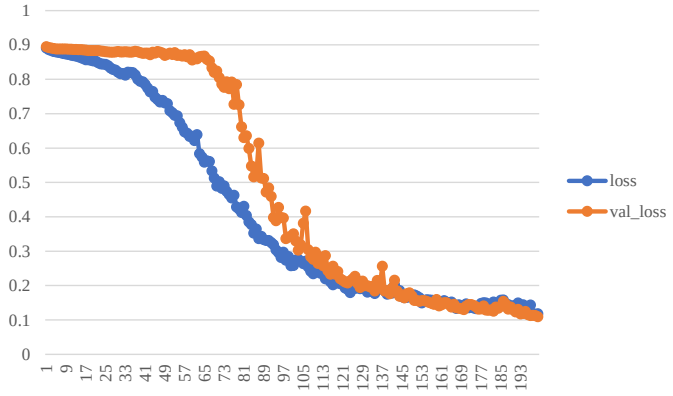
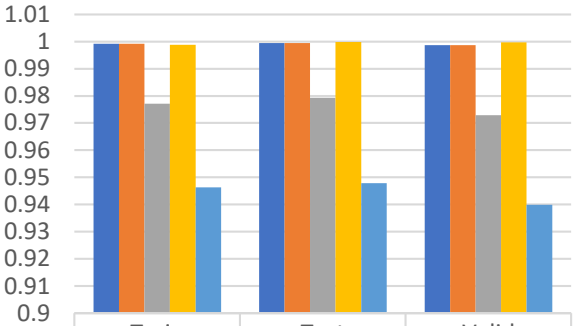


Figure 16. Focal+dice loss

The average value of 0.974184 for the AUC (Area Under the Curve) indicates that the algorithm has a high capacity to discriminate among haemorrhages and non-hemorrhages. Higher accuracy is indicated by a larger AUC, with numbers near 1 suggesting good discriminating. Consistent achievement with regard to of AUC is indicated by a small variance of 0.008255. The AUC values of 0.956402, 0.990752, and 0.956402 respectively show the range of variance in the model's capacity to discriminate between various samples. In addition, the algorithm's ability to precisely detect non-hemorrhage pixel is demonstrated by its average specific of 0.999812, which adds to its total accuracy. The model's capacity to correctly detect haemorrhage pixels is shown by its average sensibility, which is 0.939016.



	Train	Test	Valid
Accuracy	0.999244	0.999469	0.998732
Binary Accuracy	0.999244	0.999469	0.998732
AUC	0.977122	0.979326	0.972912
Specificity	0.998805	0.999862	0.999685
Sensitivity	0.946274	0.947909	0.939895

Figure 17. UHI model performance

For assessing the algorithm's effectiveness in accurately identifying haemorrhages, both metrics are essential. The ranges of quartiles (25%, 50%, and 75%) give information about the metrics' dispersion and show the variation of the model's effectiveness in various samples. The algorithm's excellent precision, good discrimination capacity, and even effectiveness in properly recognising haemorrhage and non-hemorrhage pixels are all demonstrated by the statistical evaluation of the haemorrhages dataset. The findings show how well the algorithm performs haemorrhage detection and segmentation in ocular pictures, which can be useful for identifying and tracking retinal disorders.

The initial picture in this illustration is a macular the fundus picture with retinopathy caused by diabetes symptoms. Numerous anomalies, such as exudates and microaneurysms, are visible in the picture. Specialists create the appropriate mask, which shows the areas that are important where these aberrations are found. The mask assists in emphasising the regions that require focus for additional investigation and evaluation. The network processes the picture for classification utilising the Dense-ED-UHI: Encoder Decoder based U-Net Hybrid Inception model. Intricate characteristics may be captured and accurate segmentation operations can be carried out using the U-Net Fusion framework. The aneurysm and exudate zones are correctly identified and highlighted on the segments map produced by U-Net Inception, which closely matches the surface truth mask. Figure 18 and 19 illustrates the prediction of U-Net Inception and Resnet Models.

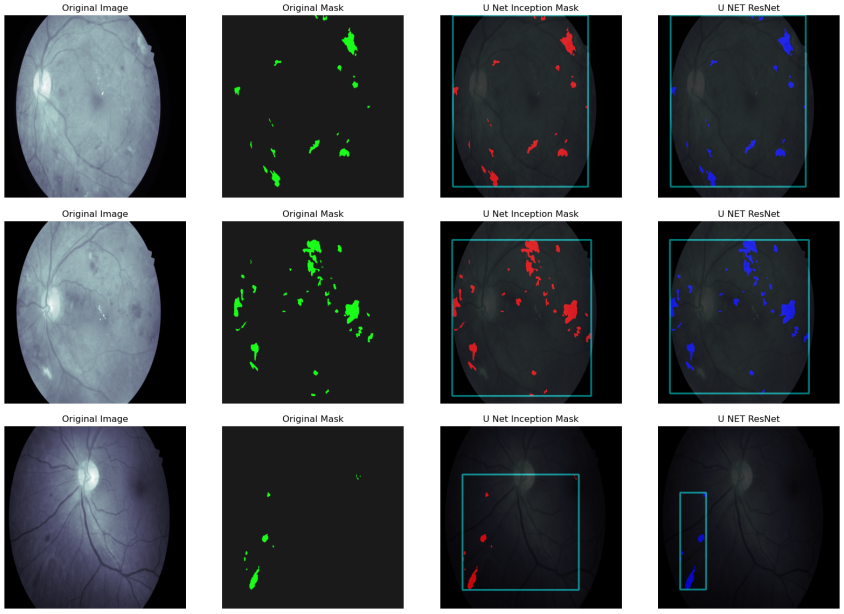


Figure 18. Mask prediction and BBox using Dense-ED-UHI: encoder decoder based unet hybrid inception (proposed model)

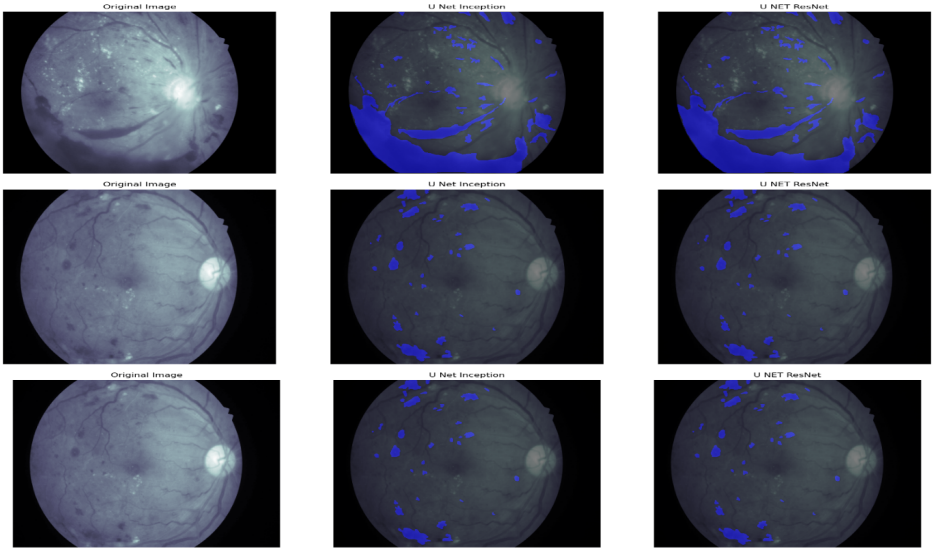


Figure 19. Comparison with other state of the art backbone (Resent50)

4.3.3. Microaneurysms

Figure 20, Figure 21 and Figure 22 describe the training and UHI network on Microaneurysms dataset. As per Table 10 and Figure 22 the algorithm has an excellent degree of general precision in detecting microaneurysms, as indicated by a median accuracy of 0.999264. It indicates that a large number of pixels in the picture, comprising all genuine positives and true negatives, are classified properly by the model.

Accuracy in Binary: the accuracy in basic measure similarly exhibits a high average value of 0.999264. This statistic, which solely takes into account true positives and true negatives, assesses the precision with which microaneurysms are classified. The elevated score shows how well the representation can separate microaneurysms from other areas of the image.

Table 10
Statistical analysis

Statistical test	Accuracy	Binary accuracy	AUC	Specificty	Sensitivity
mean	0.999264	0.999264	0.974184	0.999812	0.939016
std	0.000230	0.000230	0.008255	0.000072	0.018514
min	0.998607	0.998607	0.956402	0.999574	0.901336
25%	0.999184	0.999184	0.968939	0.999759	0.927588
50%	0.999331	0.999331	0.973982	0.999836	0.938238
75%	0.999403	0.999403	0.979575	0.999862	0.948270
max	0.999657	0.999657	0.990752	0.999934	0.976834

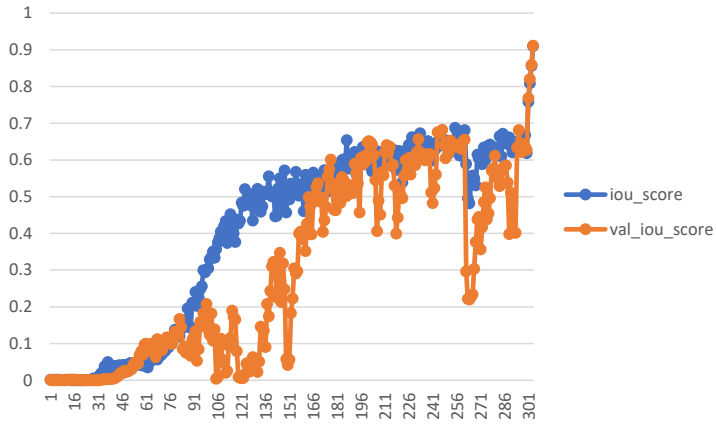


Figure 20. Training IOU_score

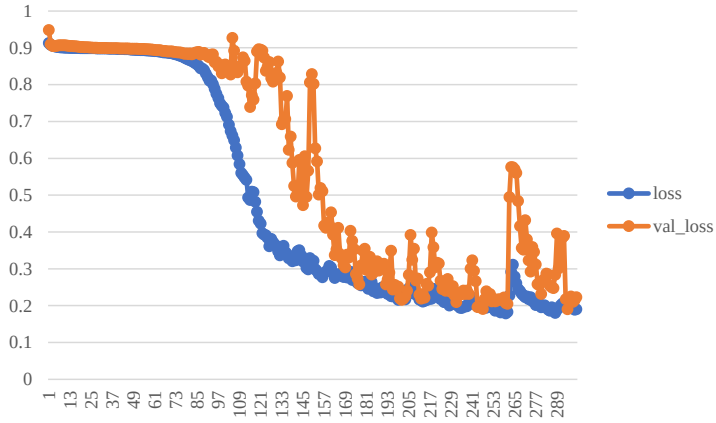
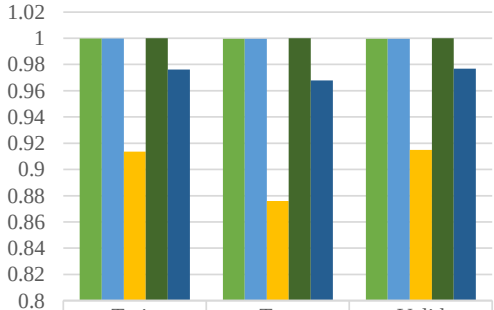


Figure 21. Focal+dice loss

The system's capacity to accurately detect non-microaneurysm areas in the picture is demonstrated by its average specificity of 0.999812 for the dataset. This means that the algorithm has a small percentage of fake positives and an elevated genuine negative rate. AUC (Area Under the Curve): the region underneath the Receiver Operating Characteristics (ROC) curve is represented by the AUC value of 0.974184. It gives an indication of how well the simulation can distinguish among microaneurysm-prone and non-prone locations. The statistical model performs better at differentiating among the two groups based on the area under the curve (AUC).



	Train	Test	Valid
Accuracy	0.99974	0.999681	0.99964
Binary Accuracy	0.99974	0.999681	0.99964
AUC	0.913566	0.875894	0.914899
Specificity	0.99997	0.999965	0.999958
Sensitivity	0.9761484	0.9677475	0.9768875

Figure 22. UHI model performance

A median rating of 0.939016 indicates how well the algorithm can identify microaneurysms. It displays the real negative percentage, which shows how well the algorithm can detect all of the microaneurysms visible in the picture. According to the statistical results, the algorithm does a good job of detecting microaneurysms generally. The model's efficiency in reliably classifying microaneurysm areas is demonstrated by its outstanding accuracy, dichotomous accuracy, and AUC scores. Furthermore, the model's excellent sensitivity and specificity scores show that it can reliably detect most aneurysm and distinguish non-microaneurysm locations. It's significant to keep in mind that the usual deviation numbers reveal information about the model's efficacy and inconsistency across different examples or datasets. The accuracy, categorical precision, AUC, particularity, and sensitivities have relatively small variances, which shows that the algorithm performs consistently throughout the assessed data.

In conclusion, the statistical evaluation shows that the algorithm consistently performs well throughout various data sets and reaches a high level of precision and efficacy in recognising microaneurysms. These findings point to the algorithm's possibility of helping doctors identify and diagnose microaneurysms promptly, and this is essential for treating patients with diabetes along with different ocular illnesses. Figure 23 and Figure 24 illustrates the prediction of U-Net Inception and Resnet Models.

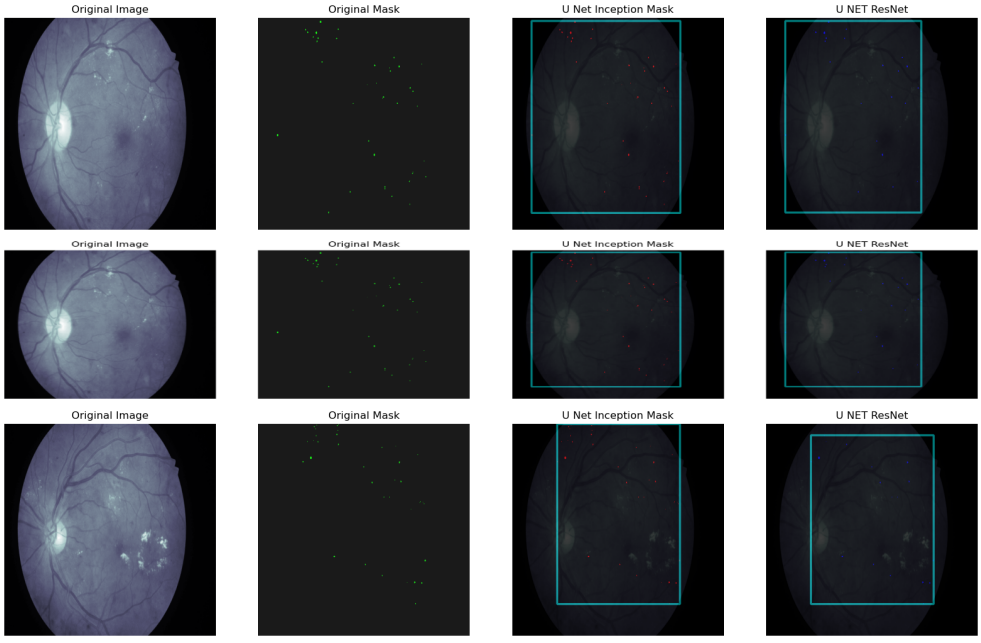


Figure 23. Mask prediction and BBox using Dense-ED-UHI: encoder decoder based U-Net hybrid inception (proposed model)

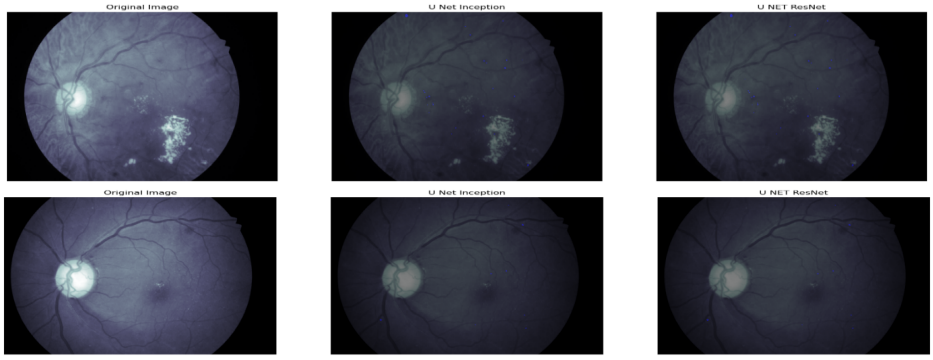


Figure 24. Comparison with other state of the art backbone (Resent50)

4.3.4. Soft exudates

The training and UHI networks on the hemorrhages dataset are shown in Figure 25, and Figure 26. The method has an outstanding level of general precision in recognizing soft exudates, as shown by a median accuracy of 0.999521, as shown in Table 11 and Figure 27. The efficiency of the Hybrid U-Net Inception model on the Soft Exudates IRID datasets seems nothing short of remarkable. The algorithm demonstrates its exceptional capacity to precisely categorise soft exudates with absolute accuracy scores of roughly 99.95% on the initial training, testing, and validation sets. Additionally, the AUC values are excellent, especially on the test and validation sets, where they are 99.90% and 99.99%, respectively, demonstrating the model's great ability to distinguish between positive and negative instances. In order to reduce the number of false positives in medical imaging analysis, the model also excels in specificity, scoring over 99.97% on all datasets. Furthermore metrics, sensitivity – which measures the model's capacity to identify positive cases – remains quite high, particularly for the test and validation sets, where it is 96.08% and 96.31%, respectively. In clinical settings, when avoiding false positives is crucial, this minor trade-off between specificity and sensitivity may be acceptable. After further validation and testing, the Hybrid U-Net Inception model exhibits strong and encouraging performance, demonstrating its potential value in real-world clinical applications.

Table 11
Statistical analysis

Statistical test	Accuracy	Binary accuracy	AUC	Specificty	Sensitivity
mean	0.999264	0.999264	0.974184	0.999812	0.939016
std	0.000230	0.000230	0.008255	0.000072	0.018514
min	0.998607	0.998607	0.956402	0.999574	0.901336
25%	0.999184	0.999184	0.968939	0.999759	0.927588
50%	0.999331	0.999331	0.973982	0.999836	0.938238
75%	0.999403	0.999403	0.979575	0.999862	0.948270
max	0.999657	0.999657	0.990752	0.999934	0.976834

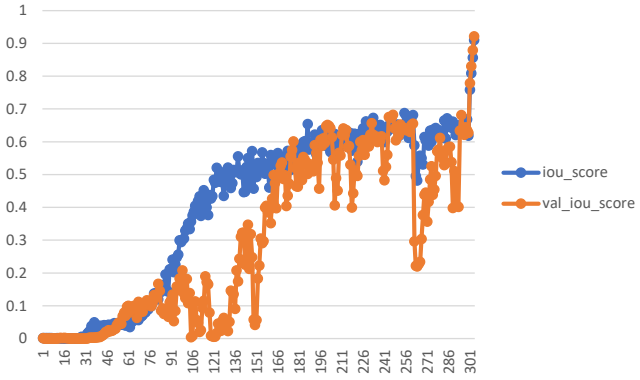


Figure 25. Training IOU_score

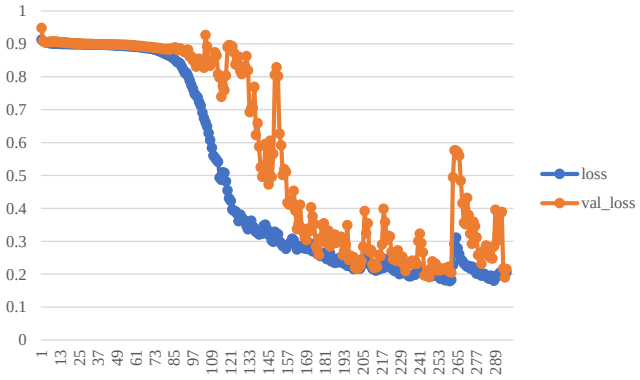
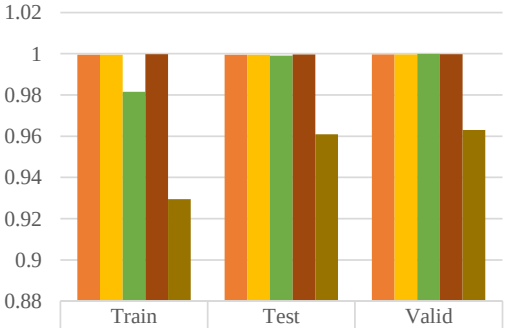


Figure 26. Focal+Dice Loss



	Train	Test	Valid
Accuracy	0.999545	0.999485	0.999671
Binary Accuracy	0.999545	0.999485	0.999671
AUC	0.981519	0.999047	0.999875
Specificity	0.999788	0.999668	0.999818
Sensitivity	0.929511	0.960885	0.963076

Figure 27. UHI model performance

Figures 28 and Figure 29 show how the Unet Inception and Resnet Models predict.

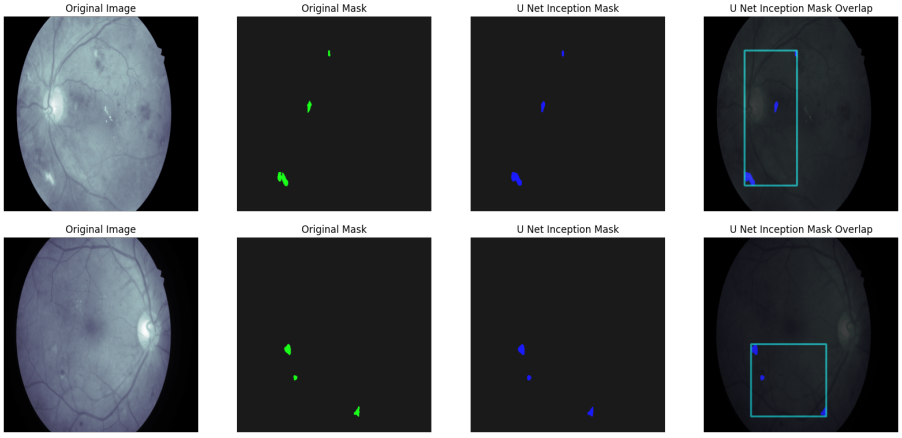


Figure 28. Mask prediction and BBox using Dense-ED-UHI: encoder decoder based U-Net hybrid inception (proposed model)

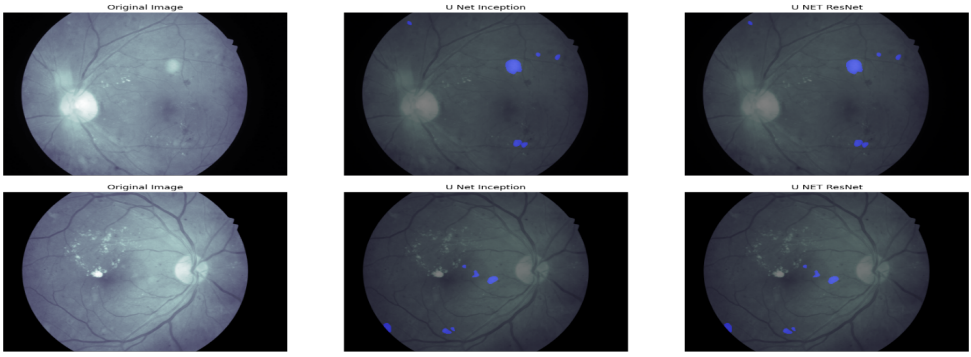


Figure 29. Comparison with other state of the art backbone (Resent50)

4.4. Comparative analysis

The comparison Table 12 illustrates how the Dense-ED-UHI model fared better than other evaluated models when specificity and accuracy were taken into account.

The model put out by [21] outperformed earlier research, achieving a sensitivity of 80.32% and a specificity of 99.83% across many datasets. In their analysis on the e-ophtha dataset for exudates identification, Chudzik et al. [9] found that specificities were higher than 99.97% and sensitivities ranged from 84.58% to 86.66%. Li et al.'s [23] work on the DDR dataset revealed an accuracy of 82.84%, whilst Playout et al.'s [32] research attained a sensitivity of 80.02% and a specificity of 82%.

Xia et al. [44], in comparison, obtained 82% sensitivity and 98.3% specificity on the CHASE dataset and 81.2% sensitivity and 98% specificity on the DRIVE dataset. Sensitivities and specificities were found to be between 80 and 83% in Oh et al.'s [27] research using ETDRS datasets.

Table 12
Comparative study

Study	Dataset	Sensitivity [%]	Specificity [%]	Accuracy [%]
[21]	HEIMED, e-ophtha, DIARETDB	80.32	99.83	–
[9]	e-ophtha (exudates)	86.66	99.98	–
[10]	e-ophtha (exudates)	84.58	99.97	–
[32]	e-ophtha (exudates)	80.02	–	82
[23]	DDR	–	–	82.84
[44]	DRIVE	81.2	98	95.4
	CHASE	82	98.3	97
[27]	ETDRS 7SF	83.38	83.41	83.3
	ETDRS F1-F2	80.60	80.61	80.6

5. Discussion

The presented research focuses on addressing the challenges of diabetic retinopathy (DR) through a novel approach employing a combinative method utilizing U-Net with a modified Inception architecture. One of the leading causes of visual impairment worldwide is diabetic retinopathy, for which prompt identification is essential to successful treatment. Deep neural architecture, notably encoder-decoder modelling using convolutional architectures like Inception and Residual Connection, is included into the suggested model. To improve performance in image processing tasks like semantic segmentation or image-to-image translation, the U-Net and Inception architectures are combined. Prominent for its effectiveness in semantic segmentation, the encoder-decoder structure of the U-Net architecture allows for accurate localization. Conversely, GoogLeNet's Inception design, which minimises computational cost by using convolutional filters of different widths inside a single layer, is excellent at collecting multi-scale information. By combining these designs, Inception modules are added to the U-Net encoder (3 and 4), improving feature extraction at various scales. Similar changes are made to the decoder, where layers of upsampling combine with feature concatenation from scale-matched encoder features. Most importantly, skip connection retention guarantees gradients and spatial information are preserved during training. The last layers are customized for each job; for pixel-wise predictions, they usually consist of convolutional layers followed by sigmoid or softmax

activations. This combination combines the multi-scale feature extraction power of Inception with the accurate localization expertise of U-Net to provide better results, especially in applications where collecting data at several sizes is critical.

The efficiency of the modified Inception deep feature extractor is shown by the study's validation on the IDRid APTOS 2019 contest dataset, which produced an exceptional classification accuracy of 94.21% across classes, outperforming the comparison with Resnet. The suggested model attains a segmentation test accuracy of 99.90% on the IDRid dataset across various classes, surpassing Resnet. Specifically, it achieves testing accuracies of 92.62% for retinopathy grade classification and 95.80% for DME classification. Moreover, the model demonstrates mean accuracies of 99.89%, 99.84%, 99.96%, and 99.97% for Hard Exudates, hemorrhage, Microaneurysms, and Soft Exudates respectively. A hybrid dense-ED-UHI model, an encoder-decoder-based U-Net with an inception architecture, is used in the article. It is cross-validated 15 times across four classes of IDRID dataset.

The model performs very well; the haemorrhages dataset examination confirms this, with mean accuracy and binary accuracy reaching 99.84%. The model is successful in identifying true positives and negatives, as seen by the AUC and sensitivity values, which highlight the model's accuracy in differentiating between advantageous and harmful scenarios. In the context of DR classification systems, the research emphasizes the relevance of the custom U-Net model with Inception architecture and a novel up-sampling technique employing pixel-wise periodic shuffling convolution. The comprehensive validation of the proposed framework against extant literature presents it as a promising advancement that might lead to better outcomes and therapies for people with visual impairment worldwide. When it comes to particularity and accuracy, the Dense-ED-UHI model performs better than the other models that were assessed. The dataset's ability to properly identify and distinguish retinal disorders is supported by its excellent accuracy scores. Because of the model's proven specificity, false positives are less likely to occur and unfavourable instances are accurately recognised. Together, these results highlight the Dense-ED-UHI model's effectiveness and dependability, highlighting its potential for accurately classifying retinal disorders.

Lastly, with regard to microaneurysms, the mathematical models show remarkable precision and binary accuracy. With a 99.96% median accuracy, AUC values, sensitivity values, and correct classification of unfavourable situations, the model demonstrates its ability to discriminate between advantageous and harmful cases while effectively identifying true positives. The high sensitivity values demonstrate even more how well the model categorises unfavourable situations. The study report concludes by presenting a unique strategy for treating diabetic retinopathy using an advanced combinative technique. The suggested approach, verified on many datasets, has exceptional precision and efficacy in categorising retinal disorders, presenting encouraging opportunities for enhanced identification and management in the worldwide context of visual impairment.

6. Conclusion

Although one is caused by ageing and the other by varied reasons, diabetic retinopathy is complicated, multifaceted diseases, and our understanding of these diseases is continually changing. Our systems of categorising these have undergone numerous revisions over time, and they have had to be revised and updated to keep pace with medical knowledge and technological developments. A new classification system must consider the tremendous advancements made in the last few decades in disease pathophysiology, imaging technologies, artificial intelligence, and treatment. In the training set as well as the test set, the accuracy of the model was 92.5%, demonstrating its capacity to make accurate assumptions about unobserved data. At 99.66% on the test set, the level of sensitivity, or real positive rate, is also incredibly high. The simulations exhibit remarkable accuracy for the examination of hard exudates, with an average precision of 99.93%. The sensitiveness scores, with a mean sensitivity of 92.94%, show how well the models can detect true positives. The models' high ability to differentiate among both positive and negative situations is reflected in the AUC values, which range from 0.9426 to 0.9827. The precise classification of adverse cases by the models is highlighted by their elevated specific values, which range from 0.9996 to 0.9999. The data analysis indicates that the models are effective in recognising and categorising exudates that are hard overall.

The hypotheses performed exceptionally well in the evaluation of the haemorrhages dataset, with a mean accuracy and binary accuracy of 99.93%. The AUC values, which range from 0.9564 to 0.9908, show how well the models can distinguish between both benign and detrimental cases. The algorithms' extremely sensitive values, which range from 0.9013 to 0.9768, indicate their accuracy in detecting genuine positives. The precise classification of negative cases by the models is highlighted by their elevated sensitivity values, which range from 0.9996 to 0.9999. These findings show that the algorithms are capable of identifying and categorising haemorrhages The new DR classification systems make use of a customised UHI or U-Net model to conceptually segment using Inception as a spatial up-sampling method that uses pixel-wise periodic shuffling convolution. It has been thoroughly validated by comparison to the body of literature and should lead to better treatments and results for the millions of individuals who suffer from visual loss globally. In terms of correctness and accuracy, the Dense-ED-UHI: Encoder Decoder based Unet Hybrid Inception models scored better than other models that were evaluated. High accuracy ratings were attained, demonstrating the datasets' capacity to appropriately diagnose and separate the retinal diseases. The algorithm furthermore showed good specificity, demonstrating its capacity to accurately recognise the adverse cases (non-pathological areas) and prevent false positives. These findings demonstrate the efficiency and dependability of the Dense-ED-UHI: Encoder Decoder based Unet Hybrid Inception framework for segmenting retinal disease.

The mathematical models are highly accurate and binary accurate when it comes to microaneurysms, with a median accuracy of 99.93%. The AUC numbers, which

range from 0.9564 to 0.9908, show how well the models can distinguish among both beneficial and detrimental cases. The sensitive values, which range from 0.9013 to 0.9768, show how well the models are able to detect true positives. The precise classification of unfavourable cases by the models is highlighted by their elevated sensitivity values, which range from 0.9996 to 0.9999. The Dense-ED-UHI: Encoder Decoder based U-Net Hybrid Inception structure's ability to use the structure known as Inception and gather global as well as local pigment data within the image which constitutes one of its main advantages. This enables the model to effectively record fine details in retina pictures by utilising multi-scale characteristics.

In conclusion, segmenting retinal disease can greatly benefit from the application of deep learning models, particularly the UHI system and the possible integration of the Classification network for broad category classification. These simulators have proven to be highly accurate, durable, and capable of automating and assisting in the early identification and recognition of retinal disorders. Continuing this field's study and development could result in better healthcare.

Authors contribution

AB and VM created the idea for the study. AB led the writing of this paper. AB analyzed and interpreted the data. AB contributed to the implementation of the paper as well as to the identification of theoretical problems. AB contributed to the writing of the paper as well as participated in revising this manuscript. VM encouraged AB to investigate and supervised the findings of this work. All authors contributed significantly to the writing of the paper, discussed the results, reviewed the final draft, and approved the final version of this manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data availability

The data URL supporting this article has shown in Table 13.

Table 13
Data availability statement

Dataset name	Availability of data	URL
IDRID	Openly accessible data stored in a public database.	The data that support the findings of this study are available in Grand-Challenge at https://idrid.grand-challenge.org.

Conflict of interest

No authors have disclosed any conflicts of interest. The authors affirm that they have no known financial or interpersonal conflicts that would have appeared to have an impact on the research presented in this study.

References

[1] Abràmoff M.D., Lavin P.T., Birch M., Shah N., Folk J.C.: Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices, *npj Digital Medicine*, vol. 1(1), 39, 2018. doi: 10.1038/s41746-018-0040-6.

[2] Akram M., Khalid S., Khan S.: Identification and classification of microaneurysms for early detection of diabetic retinopathy, *Pattern Recognition*, vol. 46(1), pp. 107–116, 2013. doi: 10.1016/j.patcog.2012.07.002.

[3] Azzopardi G., Strisciuglio N., Vento M., Petkov N.: Trainable COSFIRE filters for vessel delineation with application to retinal images, *Medical Image Analysis*, vol. 19(1), pp. 46–57, 2015. doi: 10.1016/j.media.2014.08.002.

[4] Bali A., Mansotra V.: Deep Learning-based Techniques for the Automatic Classification of Fundus Images: A Comparative Study. In: *2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, pp. 351–359, 2021. doi: 10.1109/ICAC3N53548.2021.9725464.

[5] Bali A., Mansotra V.: An Overview of Retinal Image Classification-Techniques and Challenges. In: *2021 First International Conference on Advances in Computing and Future Communication Technologies (ICACFCT)*, pp. 91–97, 2021. doi: 10.1109/ICACFCT53978.2021.9837371.

[6] Bali A., Mansotra V.: FUNDUS and OCT Image Classification Using DL Techniques. In: V.S. Rathore, S.C. Sharma, J.M.R.S. Tavares, C. Moreira, B. Surendiran (eds.), *Rising Threats in Expert Applications and Solutions. Proceedings of FICR-TEAS 2022*, Lecture Notes in Networks and Systems, vol. 434, Springer, 2022. doi: 10.1007/978-981-19-1122-4_8.

[7] Bali A., Mansotra V.: Analysis of Deep Learning Techniques for Prediction of Eye Diseases: A Systematic Review, *Archives of Computational Methods in Engineering*, vol. 31, pp. 487–520, 2024. doi: 10.1007/s11831-023-09989-8.

[8] Bali A., Mansotra V.: Multiclass multilabel ophthalmological fundus image classification based on optimised deep feature space evolutionary model, *Multimedia Tools and Applicatios*, vol. 83, pp. 49813–49843, 2024. doi: 10.1007/s11042-023-17530-z.

[9] Chudzik P., Al-Diri B., Caliva F., Ometto G., Hunter A.: Exudates segmentation using fully convolutional neural network and auxiliary codebook. In: *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 770–773, 2018. doi: 10.1109/embc.2018.8512354.

[10] Chudzik P., Majumdar S., Caliva F., Al-Diri B., Hunter A.: Exudate segmentation using fully convolutional neural networks and inception modules. In: E.D. Angelini, B.A. Landman (eds.), *Medical Imaging 2018: Image Processing*, vol. 10574, 1057430, International Society for Optics and Photonics, SPIE, 2018. doi: 10.1117/12.2293549.

[11] Di Cesare M.: Global trends of chronic non-communicable diseases risk factors, *European Journal of Public Health*, vol. 29(Supplement_4), ckz185-196, 2019. doi: 10.1093/eurpub/ckz185.196.

[12] Early Treatment Diabetic Retinopathy Study Research Group: Fundus photographic risk factors for progression of diabetic retinopathy: ETDRS report number 12, *Ophthalmology*, vol. 98(5), pp. 823–833, 1991.

[13] Franklin S.W., Rajan S.E.: Computerized screening of diabetic retinopathy employing blood vessel segmentation in retinal images, *Biocybernetics and Biomedical Engineering*, vol. 34(2), pp. 117–124, 2014. doi: 10.1016/j.bbe.2014.01.004.

[14] Fraz M.M., Barman S.A., Remagnino P., Hoppe A., Basit A., Uyyanonvara B., Rudnicka A.R., Owen C.G.: An approach to localize the retinal blood vessels using bit planes and centerline detection, *Computer Methods and Programs in Biomedicine*, vol. 108(2), pp. 600–616, 2012. doi: 10.1016/j.cmpb.2011.08.009.

[15] Gardner G.G., Keating D., Williamson T.H., Elliott A.T.: Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool, *British Journal of Ophthalmology*, vol. 80(11), pp. 940–944, 1996. doi: 10.1136/bjo.80.11.940.

[16] Gargeya R., Leng T.: Automated identification of diabetic retinopathy using deep learning, *Ophthalmology*, vol. 124(7), pp. 962–969, 2017. doi: 10.1016/j.ophtha.2017.02.008.

[17] Gulshan V., Peng L., Coram M., Stumpe M.C., Wu D., Narayanaswamy A., Venugopalan S., *et al.*: Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, *JAMA*, vol. 316(22), pp. 2402–2410, 2016. doi: 10.1001/jama.2016.17216.

[18] Guo S., Li T., Kang H., Li N., Zhang Y., Wang K.: L-Seg: An end-to-end unified framework for multi-lesion segmentation of fundus images, *Neurocomputing*, vol. 349, pp. 52–63, 2019. doi: 10.1016/j.neucom.2019.04.019.

[19] Hagos M.T., Kant S.: Transfer learning based detection of diabetic retinopathy from small dataset, *arXiv preprint arXiv:190507203*, 2019.

[20] Haldorai A., Ramu A., Chow C.O.: Editorial: Big Data Innovation for Sustainable Cognitive Computing, *Mobile Networks and Applications*, vol. 24, pp. 221–223, 2019. doi: 10.1007/s11036-018-1198-5.

[21] Imani E., Pourreza H.R.: A novel method for retinal exudate segmentation using signal separation algorithm, *Computer Methods and Programs in Biomedicine*, vol. 133, pp. 195–205, 2016. doi: 10.1016/j.cmpb.2016.05.016.

[22] Kassani S.H., Kassani P.H., Khazaeinezhad R., Wesolowski M.J., Schneider K.A., Deters R.: Diabetic Retinopathy Classification Using a Modified Xception Architecture. In: *2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 1–6, IEEE, 2019. doi: 10.1109/isspit47144.2019.9001846.

[23] Li T., Gao Y., Wang K., Guo S., Liu H., Kang H.: Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening, *Information Sciences*, vol. 501, pp. 511–522, 2019. doi: 10.1016/j.ins.2019.06.011.

[24] Lian S., Li L., Lian G., Xiao X., Luo Z., Li S.: A Global and Local Enhanced Residual U-Net for Accurate Retinal Vessel Segmentation, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 18(3), pp. 852–862, 2019. doi: 10.1109/tcbb.2019.2917188.

[25] Mansour R.F.: Deep-learning-based automatic computer-aided diagnosis system for diabetic retinopathy, *Biomedical Engineering Letters*, vol. 8, pp. 41–57, 2018.

[26] Noh H., Hong S., Han B.: Learning deconvolution network for semantic segmentation. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1520–1528, 2015. doi: 10.1109/iccv.2015.178.

[27] Oh K., Kang H.M., Leem D., Lee H., Seo K.Y.: Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images, *Scientific Reports*, 2021. doi: 10.1038/s41598-021-81539-3.

[28] Orlando J., Prokofyeva E., Del Fresno M., Blaschko M.: An ensemble deep learning based approach for red lesion detection in fundus images, *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 115–127, 2018. doi: 10.1016/j.cmpb.2017.10.017.

[29] Orlando J.I., Prokofyeva E., Del Fresno M., Blaschko M.B.: An ensemble deep learning based approach for red lesion detection in fundus images, *Computer Methods and Programs in Biomedicine*, vol. 153, pp. 115–127, 2018. doi: 10.1016/j.cmpb.2017.10.017.

[30] Osareh A., Mirmehdi M., Thomas B., Markham R.: Automated identification of diabetic retinal exudates in digital colour images, *The British Journal of Ophthalmology*, vol. 87(10), pp. 1220–1223, 2003. doi: 10.1136/bjo.87.10.1220.

[31] Partovi M., Rasta S.H., Javadzadeh A.: Automatic detection of retinal exudates in fundus images of diabetic retinopathy patients, *Journal of Research in Clinical Medicine*, vol. 4(2), pp. 104–109, 2016. doi: 10.15171/jarcm.2016.017.

[32] Playout C., Duval R., Cheriet F.: A novel weakly supervised multitask architecture for retinal lesions segmentation on fundus images, *IEEE Transactions on Medical Imaging*, vol. 38(10), pp. 2434–2444, 2019. doi: 10.1109/tmi.2019.2906319.

[33] Quellec G., Charrière K., Boudi Y., Cochener B., Lamard M.: Deep image mining for diabetic retinopathy screening, *Medical Image Analysis*, vol. 39, pp. 178–193, 2017. doi: 10.1016/j.media.2017.04.012.

[34] Qureshi I., Ma J., Abbas Q.: Diabetic retinopathy detection and stage classification in eye fundus images using active deep learning, *Multimedia Tools and Applications*, vol. 80, pp. 11691–11721, 2021. doi: 10.1007/s11042-020-10238-4.

[35] Romero-Aroca P., Verges-Puig R., de la Torre J., Valls A., Relaño-Barambio N., Puig D., Baget-Bernaldiz M.: Validation of a deep learning algorithm for diabetic retinopathy, *Telemedicine and e-Health*, vol. 26(8), pp. 1001–1009, 2020. doi: 10.1089/tmj.2019.0137.

[36] Roychowdhury S., Koozekanani D.D., Parhi K.K.: DREAM: diabetic retinopathy analysis using machine learning, *IEEE Journal of Biomedical and Health Informatics*, vol. 18(5), pp. 1717–1728, 2013. doi: 10.1109/jbhi.2013.2294635.

[37] Shanthi T., Sabeenian R.S.: Modified Alexnet architecture for classification of diabetic retinopathy images, *Computers & Electrical Engineering*, vol. 76, pp. 56–64, 2019. doi: 10.1016/j.compeleceng.2019.03.004.

[38] Shen W., Zhou M., Yang F., Yu D., Dong D., Yang C., Zang Y., Tian J.: Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification, *Pattern Recognition*, vol. 61, pp. 663–673, 2017. doi: 10.1016/j.patcog.2016.05.029.

[39] Sinthanayothin C., Boyce J.F., Williamson T.H., Cook H.L., Mensah E., Lal S., Usher D.: Automated detection of diabetic retinopathy on digital fundus images, *Diabetic Medicine*, vol. 19(2), pp. 105–112, 2002. doi: 10.1046/j.1464-5491.2002.00613.x.

[40] Soomro T.A., Gao J., Khan T., Hani A.F.M., Khan M.A., Paul M.: Computerised approaches for the detection of diabetic retinopathy using retinal fundus images: a survey, *Pattern Analysis and Applications*, vol. 20, pp. 927–961, 2017. doi: 10.1007/s10044-017-0630-y.

[41] Sopharak A., Uyyanonvara B., Barman S., Williamson T.H.: Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods, *Computerized Medical Imaging and Graphics*, vol. 32(8), pp. 720–727, 2008. doi: 10.1016/j.compmedimag.2008.08.009.

[42] Ting D.S.W., Tan G.S.W., Agrawal R., Yanagi Y., Sie N.M., Wong C.W., San Yeo I.Y., Lee S.Y., Cheung C.M.G., Wong T.Y.: Optical coherence tomographic angiography in type 2 diabetes and diabetic retinopathy, *JAMA Ophthalmology*, vol. 135(4), pp. 306–312, 2017. doi: 10.1001/jamaophthalmol.2016.5877.

[43] Walter T., Massin P., Erginay A., Ordonez R., Jeulin C., Klein J.C.: Automatic detection of microaneurysms in color fundus images, *Medical Image Analysis*, vol. 11(6), pp. 555–566, 2007. doi: 10.1016/j.media.2007.05.001.

[44] Xia Z., Liu Q., Zhang Z., Wang Y.: Segmentation of exudates in fundus images using fully convolutional neural networks, *Journal of Medical Imaging and Health Informatics*, 2020.

[45] Xiancheng W., Wei L., Bingyi M., He J., Jiang Z., Xu W., Ji Z., Hong G., Zhaomeng S.: Retina blood vessel segmentation using a U-Net based Convolutional neural network. In: *Procedia Computer Science: International Conference on Data Science (ICDS 2018)*, pp. 8–9, 2018.

[46] Xu X., Tan T., Xu F.: An improved U-Net architecture for simultaneous arteriole and venule segmentation in fundus image. In: M. Nixon, S. Mahmoodi, R. Zwiggelaar (eds.), *Medical Image Understanding and Analysis: 22nd Conference, MIUA 2018, Southampton, UK, July 9–11, 2018, Proceedings*, pp. 333–340, Springer International Publishing, 2018. doi: 10.1007/978-3-319-95921-4_31.

[47] Yan Z., Han X., Wang C., Qiu Y., Xiong Z., Cui S.: Learning mutually local-global u-nets for high-resolution retinal lesion segmentation in fundus images. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 597–600, IEEE, 2019. doi: 10.1109/isbi.2019.8759579.

[48] Zhang W., Zhong J., Yang S., Gao Z., Hu J., Chen Y., Yi Z.: Automated identification and grading system of diabetic retinopathy using deep neural networks, *Knowledge-Based Systems*, vol. 175, pp. 12–25, 2019. doi: 10.1016/j.knosys.2019.03.016.

[49] Zhang X., Thibault G., Decencière E., Marcotegui B., Laû B., Danno R., Cazuguel G., *et al.*: Exudate detection in color retinal images for mass screening of diabetic retinopathy, *Medical Image Analysis*, vol. 18(7), pp. 1026–1043, 2014. doi: 10.1016/j.media.2014.05.004.

[50] Zhao Y.Q., Wang X.H., Wang X.F., Shih F.Y.: Retinal vessels segmentation based on level set and region growing, *Pattern Recognition*, vol. 47(7), pp. 2437–2446, 2014. doi: 10.1016/j.patcog.2014.01.006.

Affiliations

Le Akanksha Bali
University of Jammu, Department of Computer Science and IT, akankshabali5@gmail.com

Kuljeet Singh
University of Jammu, Department of Computer Science and IT, khalsakuljeet004@gmail.com

Vibhakar Mansotra
University of Jammu, Department of Computer Science and IT, vibhakar20@yahoo.co.in

Received: 17.12.2023
Revised: 9.06.2024
Accepted: 9.06.2024

Computer Science • 25(4) 2024 https://doi.org/10.7494/csci.2024.25.4.5377

Le Danh Tai
Ta Minh Thanh

A PROPOSAL OF DIGITAL CONTENTS COPYRIGHT PROTECTION BY USING BLOCKMARKING TECHNIQUE

Abstract *Recently, Blockmarking technique [15] is proposed for a new hybrid model based on the combination of blockchain and watermarking method. In this model, it not only achieves the goal of image copyright protection but also stores the image into the blockchain network such as IPFS system. In this paper, we propose a new DRM system by inheriting the idea of Blockmarking. The copyright contents can be distributed via IPFS blockchain, then be restored by using the reconstruction license for each legal user. Also, in our method, based on the reconstruction licenses, the distributed contents can be reconstructed from IPFS with various watermarking patterns. It helps us can manage the legal users and trace the traitor if a dispute occurs. The experimental results show that our method successfully achieved the purpose of digital copyright protection.*

Keywords InterPlanetary File System (IPFS), copyrights protection, Distributed Watermarking Method, Digital Rights Management (DRM)

Citation Computer Science 25(4) 2024: 621–636

1. Introduction

1.1. Overview

The raging of piracy has brought immeasurable losses to content creators, especially prominent in areas such as news, design, photography and e-commerce. However, because the digital content itself is difficult to identify with embezzlement and the legal process takes a long time, victims often choose the actual fee of the violation, which also makes to the increase copyright infringement.

In recent years, with increasing awareness of copyright protection at national, social and individual levels, the protection of copyright such as videos, music and literature has taken a big step forward. Additionally, the content creators also create the non-fungible tokens (NFT) by minting their digital contents via blockchain marketplace. In order to create an NFT art, a creator needs to upload their digital content to a marketplace that supports NFTs, such as OpenSea[1]. The platform will then mint the content into an NFT and assign it a unique identifier and metadata. The creator can then set the price and terms of sale for their NFT, such as a fixed price, an auction, or a royalty fee. The platform will also charge a fee for the creation and listing of the NFT. To buy an NFT, a buyer needs to have a digital wallet that supports NFTs, such as MetaMask[2]. Such NFTs are very valuable. However, due to some characteristics of the digital contents itself, the process of copyright protection has progressed slowly. The content of the digital contents is indispensable, the value of the digital contents is increasingly enhanced, and the copyright protection of the digital contents becomes indispensable.

To address such digital content infringement on the market today, the researchers hope to solve the dilemma of rights protection through technology, to be able to trace the original author by how to label and engrave their own work imprints. Blind watermark is a type of digital watermarking technology, which can hide digital information in an digital contents. The processed image appears to be unchanged, but in fact the image already has a unique identifier. Regardless of whether it is cut, pasted, rotated, zoomed or added text or filters, the content of the watermark will be affected to some extent. This allows copyright protection and tracking without damaging the original work and going unnoticed.

The most robust watermarking techniques focused on copyright protection are frequency domain based watermarking methods. The frequency domain can be applied on the watermarking techniques by single frequency domain or hybrid frequency domain. Such combination of frequency domain can define the robustness of watermarking methods [5, 11, 13]. The disadvantage of this watermarking algorithm is that it is not possible to embed one bit of watermark in all blocks, resulting in low capacity. To improve payload capacity, all blocks are used for embedding purposes in [10]. The

[1] https://opensea.io/
[2] https://metamask.io/

robustness of the proposed scheme was tested against various single and combined attacks, and a good quality watermark was extracted even after multiple simultaneous attacks on the system.

In order to get the balance of imperceptibility and the robustness for frequency domain watermarking methods, Thanh *et al.* had proposed the q-logarithm frequency domain (q-LFD) for image watermarking. They have applied the q-LFD for single frequency domain in order to create a new frequency domain such as q-DCT [23], q-DWT [20], and q-SVD [21]. It is clear that depends on the values of q parameter, they could provide a new frequency domain for robustness and high quality of image watermarking methods. However, to find out the optimal values of q parameter, it's algorithm is quite complicated for analysis.

Nowadays, the non-fungible tokens (NFT) marketplaces cannot handle the challenges related to NFT ownership claims[3], illegal redistribution, and data ownership traceability[4]. The creation of NFTs has specifically led to a lucrative market for verifying ownership of unique digital assets, including digital art products. However, this NFT trading market also raises risks such as fraud, stolen works, authenticity and copyright issues. Illegal traders exploit the market by trading unauthorized copies of digital objects as NFTs.

To overcome these problems, the watermarking methods [18] can combine with blockchain technique [19] to make the efficient NFT marketplace. Saeed *et al.* [12] proposed a marketplace based on watermarking and NFT technologies. In their system, the ownership data is stored as an NFT, then the copyright information is embedded into the content of the NFT. The watermarked information can be extracted from the watermarked NFT to identify the owner and the buyer of the traded data. Dalla *et al.* [4] also proposed the digital watermarking as a means to establish the authenticity of NFTs and showed the potential of NFTs ownership protection technique. Sarad Venugopalan and Heiko Aydt [25] proposed a solution that puts control back in the hands of information owners by storing encrypted content on a data warehouse and providing additional security against hacks and zero day exploits. Content on their data warehouse is never decrypted or returned to its owner for decryption during rekeying. Their solution seems to be good for proving the ownership of NFT, however, such system requires the complicated infrastructure.

With another approarches, Tai *et al.* [15] had proposed the Blockmarking technique to embed the various watermarks information into many distributed patches of NFT stored in IPFS. They proposed the idea to extend the model of DRM system by using the distributed feature of blockchain network. Tai *et al.* [16] also extend such model on distributing NFT image system combined with watermarking technique. They could prove that their hybrid model based on watermarking algorithm with blockchain technology can work with the copyright protection of NFT image.

[3]https://fromlight2art.com/how-to-protect-your-nfts-for-artists/
[4]https://www.imatag.com/blog/digital-art-and-nft-how-to-solve-the-trust-issue

To build a transparent digital product copyright protection system, Digital Rights Management (DRM) solution for NFT marketplace is proposed to effectively manage the processing flow of digital products from manufacturers to users. The DRM solution allows producers to control what users can do with digital content such as photos, videos, logos, audio files and so on. DRM is the management of legal access to digital content. It also can restrict the use of proprietary hardware and copyrighted works. DRM technologies govern the use, modification and distribution of copyrighted digital contents. Therefore, such DRM technologies include licensing agreements and encryption.

1.2. Classification of DRM system

Based on technical solutions to classify DRM solutions, we can divide DRM into several solutions as follows:

- **Provider-based DRM system (PDRM):** focuses on the protecting of contents provider's copyright [1, 8, 24]. It is also called provider-centric DRM solutions. That means PDRM proves the copyrights of producers when he/she claims a right to his digital contents. Therefore, even if the digital contents are sold for end users, the copyright of producers/authors is still remained.
- **User-based DRM system (UDRM):** is the users-centric DRM solution. It protects the copyrights of end users after he/she bought the digital contents. UDRM is used to implement the system for embeding the copright information of legal users into the digital contents [6, 7, 17, 18]. Therefore, UDRM requires the user's information when he/she registers to buy the digital contents. The watermarked contents from UDRM may be processed under various attacks such as blur, noise addition, soften, sharpen, JPEG conversion, Rotation, Scaling, and Translation (RST) and so on. Also, the extracted watermark is required to be clear in order to judge the rights of users.
- **Hybrid model DRM system (HDRM):** is used to protect both copyrights of providers and legal users [15, 26]. In this case, HDRM requires the watermark W from the provider P and the watermark W_u from the user U. This solution is potential for DRM because of adaptation with new technology likes AI and blockchain techniques.

Three models of DRM above can be applied on real applications. However, each has several problems. Although HDRM can improve the problems of PDRM and UDRM, it is still dependent to authority judgement with saving of copyrights information in their central database. Also, only authority judgement can judge the traitors or copyright of providers/users. If such DRM systems apply on blockchain for protecting the NFT assets, the NFT marketplace can be efficiently managed. New method of the design for embedding the watermark into digital contents before minting its NFT is required. Therefore, we need to improve such problem by employing the advantages of blockchain technology combined with watermarking method.

1.3. Our contributions

Previous DRM solutions have been effectively applied in copyright management system. DRM system consists of three components such as watermarking processing, license management, and legitimate user tracking. However, that DRM system belongs to a third party, so the above three components can be controlled by the third party. Especially, recent DRM systems are hard to apply on NFT marketplace for NFT copyright protection.

In order to solve such issue, we propose a new DRM system based on redundant digital contents distributed through a blockchain network by using many patches from digital contents. We also propose a systematic information restructuring method for content distribution and verify legitimate users. Our method can manage multiple digital patches of digital images from the blockchain network. The digital content is firstly split to get multiple patches. Then, they are saved to the InterPlanetary File System (IPFS). We also prepare some watermark patterns to embed in each digital patch before saving them on IPFS.

We emphasize the following contributions in our paper:

1. *Distributed digital contents management instead of central management*
   Our system proposes a new method for digital contents management. The copyrighted digital content is divided into many patches, called NFT assets. Such patches can be overlapped or non-overlapped. Afterwards, all patches are uploaded on IPFS system. IPFS is a modular suite of protocols purpose built for the organization and movement of content-addressed data. Therefore, the saving of digital contents does not depend on the database of third party. In order to access the distributed patches (NFTs), we can manage the content identifier, or CIDs, is a label used to point to each digital patch in IPFS. Therefore, by using the distributed NFT stored via IPFS, we can manage the copyright of NFTs more conveniently.

2. *Propose watermarking method for the distributed patches before minting NFTs*
   Our system provides new watermarking method to embed the copyright information into the distributed image patches before minting NFTs on blockchain. Therefore, in our system, before becoming NFTs, the copyright information is also embedded into the digital contents. That makes our system different from other proposals.

3. *Propose new method for managing legal users by using licensing system*
   In order to manage the legal users, we propose the licensing system to control the watermark patterns for reconstructing the watermarked digital contents by using the downloaded NFTs from IPFS. Such watermark patterns are used to identify the legal users. Also, when dispute about the right to use digital products happens, we can trace the traitor and judge the copyrights of digital content.

4. *Legal users identification using new watermark patterns*
   Instead of using only one watermark logo for identifying the copyright of digital contents, we propose new approach that employs multiple watermark logos for proving the right legal users and also tracing the traitors. This is the novel idea to improve the conditional DRM system for managing legal users.

1.4. Roadmap

The remain of our paper is organized as follows. The explaination of preliminaries is described in Section II. In Section III, the components of proposed DRM method based on digital watermarking combined blockchain network are introduced, including watermark embedding and extracting processes. Our simulation results are described in Section IV. The conclusion is shown in Section V.

2. Preliminaries

To solve the problem of Digital Rights Management abusing detection to protect multimedia content, we have proposed a new watermarking scheme based on the DCT transform domain. We employ the IPFS blockchain for distributed storage of digital contents, the watermarking technique for embedding the copyright information, and distributed patches reconstruction.

2.1. Blockchain network

Blockchain is an potential technique that encompasses many technologies,e.g. cryptography, mathematics, consensus algorithms and economic models [9]. It also enhances customer service, drives end-to-end value, and increases operational efficiency. It is a secure, shared and immutable distributed ledger (database). Such a database records all of the network's transaction data into blocks. It uses a peer-to-peer (P2P) network and a consensus mechanism to solve the problem of distributed data synchronization. Therefore, it is not necessary to have a centralized trusted authority [14].

In the blockchain, block data is defined as a back-linked record in the order of blocks of transactions. Such blockchain data can be saved in a database (as a large file). Each block in the chain can be specified using the cryptographic hash algorithm SHA256[5] on the block header. The block consists of two parts, the main structure data and the header information. The main structure data records a list of transaction information across the network, while the header information includes the hash of the previous and current block, Merkle Root, timestamp, nonce, and other information.

Since all list of transactions is permanently stored in a block, if we apply blockchain technology to manage copyright, we can track all the transactions that belong to a certain digital asset. That makes it our advantage over other DRM systems because everyone can claim the copyright of digital content through the blockchain network.

[5]https://emn178.github.io/online-tools/sha256.html

2.2. Decentralized storage: IPFS

IPFS is a P2P distribution hypermedia protocol that aims to act as a universal file system for all computing devices [3]. IPFS can be considered similar to the WWW. It is like a single BitTorrent pool that exchanges digital objects in a single Github repository. IPFS combines a decentralized hash table, data exchange and a self-certifying namespace, also forming a generic Merkle architecture.

In particular, in the IPFS system, there is no single point of failure and the nodes do not need to trust each other. Based on that, distributed digital content delivery can save network bandwidth consumption [2].

IPFS is a technology especially suitable for distributing digital content over a blockchain network. We can manage the CIDs inside the DRM system to achieve certain digital content. We will show how to integrate IPFS technology with DRM system in our proposed method. Such a combination of techniques may replace third-party functionality.

3. Proposal of Blockmarking based DRM system

3.1. Overview of our system

Store blockchain-based IPFS decentralized content on the web by dividing large digital files into patches and distributing those patches across the network. Thus, if IPFS replaces the functionality of authority judgement in a DRM system, the copyrighted contents (copyrighted NFTs) can be distributed over the blockchain network. Based on that, the storage of copyrighted content does not depend on the authority judgement. Also, when someone wants to check the copyright of any content, they can obtain the content via IPFS and extract the watermark. There are two main processes that comprise our recommended approach: embedding and distribute embedded patches in IPFS storage, copyrighted contents delivery and copyright confirmation via IPFS.

3.2. Embedding and distributing embedded patches in IPFS storage

Our idea changes the normal way to store the digital contents for sale by moving the digital contents to NFTs marketplace. We do not save the copyrighted contents into the centric database of content providers. In order to take advantage of IPFS, we consider to split the original content I into many patches. Then, we embed multiple watermarks into those patches. Afterwards, the embedded patches are distributed via IPFS by minting them to NFTs via blockchain. The workflow of this stage is shown in Figure 1.

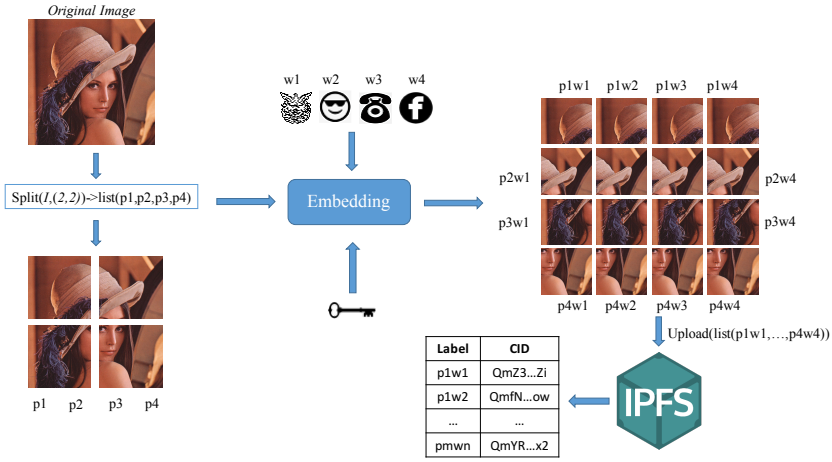


Figure 1. The process of Embedding and Distributing embedded patches in IPFS storage

The algorithm is explained as follows:

1. The original content I is split into m patches of image. Such patches denotes as $\{p1, p2, p3, \ldots, pm\}$.
2. n patterns of watermarks, *e.g.* $\{w1, w2, w3, \ldots, wn\}$ are used to embed into m patches of image. After that, we can obtain $m \times n$ watermarked patches. That means we have $\sum_{i=m}^{j=n} p_i w_j$ watermarked patches. Such watermarked patches are minted via IPFS network. After uploaded, CIDs are generated for referencing content in distributed information systems.
3. All CIDs are collected, then they are saved into key-value as {Label, CID} system.

3.3. Licensing management

After uploaded the watermarked patches (NFTs) via IPFS, we can manage all CIDs by using key-value table <Label, CID>. By using this table, we can randomly generate the license table L beforehand, then assign the license number for according legal users when he/she bought the digital contents. The detail of licensing process management is shown in Figure 2. This algorithm is introduced as follows:

1. The license table L is randomly generated based on the CIDs table. Since licenses are based on n (the number of watermark), the license number can denote as $\{L1111, L1112, L1113, \ldots, Lnnnn\}$.
2. When users want to buy the digital contents, he/she requests to provide a license number. The provider send him/her a license number from the license table L. According to such license number, n patterns of CIDs are decided. Afterwards, n patterns of image patches (NFTs) are downloaded from IPFS to combine into the watermarked contents I' for end user U.
3. After sent the license number for end user, the information of user and license number are stored into license table. Based on table L, the producer can manage the license and that of legal user.

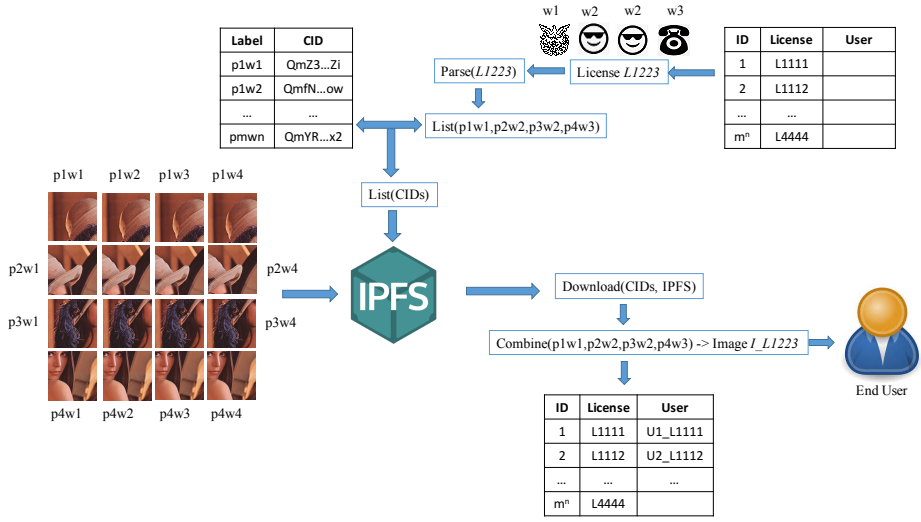


Figure 2. Licensing process management

3.4. Copyrights identification

When the copyright dispute happens or an illegal distribution of copyrighted products is discovered, it is essential to confirm the copyright and identify the legal owner. In our method, suppose there is a copyright dispute of the copyrighted content I', we need to extract the watermark patterns from I', then match the patterns via license table. According that, the information of legal users can be detected. This algorithm is shown in Figure 3.

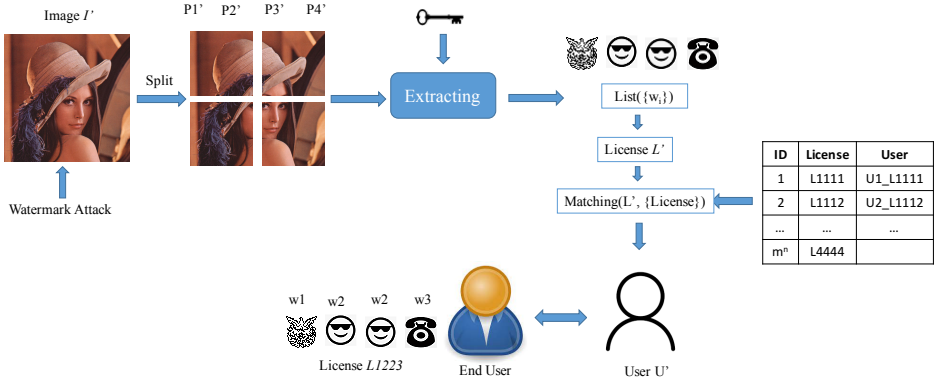


Figure 3. Copyrights identification process

1. The digital content I' is split into m patches of image. Such patches denotes as $\{p'1,' p2, p'3, \ldots, p'm\}$.
2. n patterns of watermarks, *e.g.* $\{w'1, w'2, w'3, \ldots, w'n\}$ are extracted from m patches of image. After that, we can obtain the watermark patterns. Based on the watermark patterns, we can detect the license number $Lxxx$, then we can trace the legal user of digital content.
3. After detected the legal user by using the license number $Lxxx$, we can judge the rights of watermarked contents.

According to analysis above, we emphasize that our proposed system can replace the conventional DRM system by employing the blockchain technology. We can distribute digital contents through IPFS, then anyone can check the copyright of the content transparently.

3.5. The benefit of our proposed system

In our system, the original image I is divided into m patches. Afterwards, each patch is watermarked separately with n patterns of watermark logos or ID information. In order to sell to many end users, therefore, we can obtain n^m different combinations of watermarked images employing $n \times m$ different watermarked patches. Then, $n \times m$ different watermarked patches are minted as NFTs and are saved to the IPFS network, and each patch gets unique address (CID)[6] which can be used to download the patch. The end user when buying an image obtains m CIDs that can be used to download m patches and then reconstruct the whole image with different watermark patterns.

According to above explanation of our algorithm, the main benefit of our proposed system is that it saves the disk space, computing power, and Internet network bandwidth (as some patches may be locally cached) as normally for L end users (buyers) we need to upload L different copies of watermarked (whole) images to the IPFS network. In the proposed system, the maximum number of legal users is n^m. In most cases $n = 2$ or 3 (for $m = 16$), therefore the required disk space is only doubled/tripled for managing L users with L licenses. The watermarking system might be also more immune to watermarking attacks.

4. Performance analysis and discussion

4.1. Experimental environment

In order to perform the efficiency of the proposed algorithm, we employ six color images from image database[7] with size $M \times N = 512 \times 512$ pixels. For embedding watermark patterns, we use four logomarks in our experiments, and they are the binary image with size 64×64 as shown in Figure 4.

[6] https://docs.ipfs.tech/concepts/content-addressing/#how-cids-are-created
[7] www.vision.kuee.kyoto-u.ac.jp/IUE/IMAGEDATABASE/STDIMAGES/

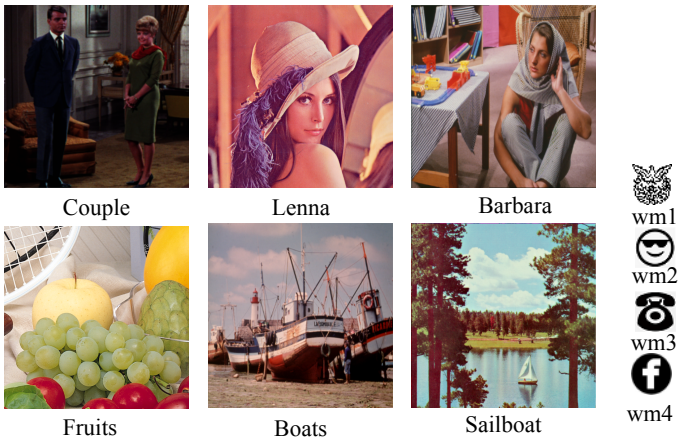


Figure 4. Experimental images

In general, to measure the efficiency of image watermarking schemes, their invisibility, robustness, and computing time are calculated and are compared each others. To evaluate the invisibility capability, we use the peak signal-to-noise ratio ($PSNR$) to measure the similarity between the original color image I and the watermarked image I' with size of $M \times N$.

The value of $PSNR$ is employed as a measure for evaluating the quality of the watermarked image comparing with that of the original image. $PSNR$ is described by the following equation:

$$PSNR = 10 \log_{10} \frac{255^2}{MSE}, \tag{1}$$

where MSE is mean square error between the original and watermarked image. MSE is defined as:

$$MSE = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I(i,j) - I'(i,j))^2 \tag{2}$$

In order to measure the robustness of our method, we also used the normalization correlation (NC) value [22] and calculated it over all extracted logomarks.

4.2. Experimental results

In order to evaluate our proposed system, we split the original image I into $n = 4$ non-overlap blocks image. We called them as Patch1, Patch2, Patch3, Patch4. Then, four logomarks patterns are embedded into each patch image for generating the embedded patches images. After that, the embedded block images are minted to NFTs to IPFS network. The testing for confirmation of image quality and the robustness of watermark extraction are performed, then the results are shown in Table 1.

Table 1
Experimental values of PSNR/NC for all images

Image	Patch1 PSNR/NC	Patch2 PSNR/NC	Patch3 PSNWNC	Patch4 PSNR/NC	PSNR Average/NC	PSNR Combine/NC
Lenna	47.21/1.0	47.63/0.9931	46.16/0.9980	48.67/0.9940	47.41/0.9963	47.24/0.9961
Couple	50.23/0.7805	50.27/0.9795	51.70/0.6543	52.25/0.8377	51.11/0.813	50.94/0.8047
Barbara	45.90/0.9958	47.79/1.0	47.04/0.9940	48.66/1.0	47.34/0.9975	47.11/0.9975
Fruits	44.96/0.9876	48.25/0.9977	47.78/0.9940	46.80/0.9940	46.95/0.9933	46.55/0.9926
Sailboat	43.98/1.0	43.62/0.9977	46.05/1.0	45.54/1.0	44.80/0.9994	44.57/0.9995
Boats	50.60/0.9979	48.84/0.9954	46.38/0.9980	46.32/0.9311	48.03/0.9806	47.41/0.9800

We tried to generate the random license, then restored the watermarked contents by retrieving the watermarked NFTs from IPFS. Afterwards, we evaluated the robustness of the embedding method by using some attacks such as Gaussian noise, salt&pepper noise, JPEG compression, erasing, change histogram, and change color. The testing of above attacks on Lena image is shown in Table 2. According Table 2, we can see that the copyright logo can be clearly recognized. The values of PSNR and NC are suitable for copyright protection. In general, the value of PSNR is over 40dB, and the value of NC is over 0.9.

Table 2
Experimental values of PSNR/NC for Lena after attacks

Attack	Patch1	Patch2	Patch3	Patch4	NC Average	NC Combine
Gaussian noise	0.7712	0.7706	0.7382	0.7194	0.7499	0.7566
Salt&Pepper noise	0.7509	0.6971	0.7618	0.7318	0.7354	0.8538
Jpeg compression	0.4015	0.4134	0.3758	0.4311	0.4055	0.4111
Erasing	0.7587	0.7293	0.7078	0.7699	0.7414	0.7276
Histogram	0.8656	0.8220	0.8223	0.8180	0.8319	0.8679
Color change	1.0	0.9931	0.9960	0.9841	0.9993	0.9950

The visualization of the attacked Lena can be seen in Figure 5. It is clear that, based on the watermarked patterns, we can match the license number and detect the legal users based on our license management system. We also retrieved the similar results based on experimental results of another images.

We also compared the proposed method with several previous methods for NFT copyright protection. Such comparison is show in Table 3. Our method and Saeed *et al.* [12] use the watermark information to prove the copyright of NFT contents. Meanwhile, Dalla *et al.* [4] only used watermark for authentication information. In case of illegal NFT redistribution, our method can track the legal users by using the license number $Lxxx$ to extract the watermark pattern. However, other methods cannot track the redistributed users when a dispute occurs.

According to above analysis, we can understand that our method can be applied for NFTs marketplace efficiently.

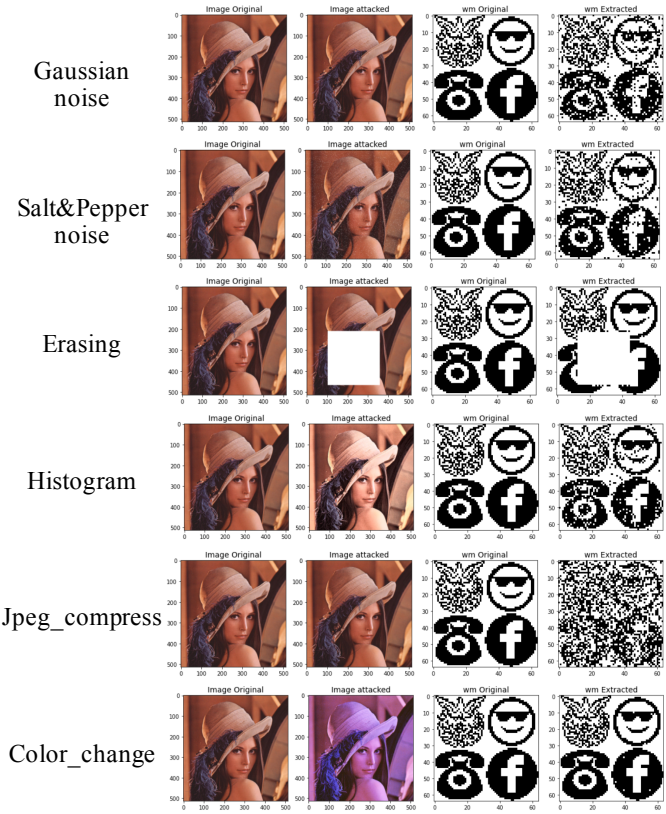


Figure 5. Logo patterns extraction under several attacks (for Lena)

Table 3
Comparison with some previous methods

Criteria	Our method	Saeed [12]	Dalla [4]
Watermarking for NFT	✓	✓	×
Legal users tracking	✓	×	×
Combination NFT patches	✓	×	×

4.3. Discussion

The limitation of the proposed method is that if the end user buys 2 or more copies of the same image with different watermarking, it is possible to construct an image with different watermarking IDs. In this case the image can be redistributed and the end users who redistribute the image may not be properly identified. That means the extracted watermarking pattern does not belong to the licenses that are generated beforehand. Therefore, the image is redistributed and the end users who redistribute

the image may not be properly identified, however, the users use such image with wrong watermark pattern that does not belong to the license database L, such users are judges as illegal users.

The solution to the above problem might be that one patch has as many watermarks as the number of users or the n is much larger than required and additional algorithm is used to track the users who redistributed the image. This solution will be solved in the future works.

5. Conclusions

We have presented a new distributed image by using IPFS combined watermarking method for copyright protection and users authentication system without third parties. The original digital images are split into many blocks image, then are embedded with various watermark patterns, and uploaded on the online storage via IPFS. The license number is used to generate the watermark patterns for achieving the embedded blocks image from IPFS. Those also can be used for specify the legal users. According to the experimental results, the watermark patterns could be successfully extracted and distinguished each others.

References

[1] Asikuzzaman M., Pickering M.R.: An Overview of Digital Video Watermarking, *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28(9), pp. 2131–2153, 2018. doi: 10.1109/TCSVT.2017.2712162.

[2] Badr B., Horrocks R., Wu X.: *Blockchain by Example: A Developer's Guide to Creating Decentralized Applications Using Bitcoin, Ethereum, and Hyperledger*, Packt Publishing, 2018.

[3] Benet J.: IPFS – Content Addressed, Versioned, P2P File System, 2014. doi: 10.48550/ARXIV.1407.3561.

[4] Dalla Preda M., Masaia F.: Exploring NFT Validation through Digital Watermarking. In: *The 18th International Conference on Availability, Reliability and Security (ARES 2023), August 29 September 01, 2023, Benevento, Italy*, Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3600160.3605063.

[5] Gaid S.K., Jabbar K.K.: Frequency Domain for Color Image Authentication Proofing, *Journal of Physics: Conference Series*, vol. 1963(1), 012094, 2021. doi: 10.1088/1742-6596/1963/1/012094.

[6] Iwakiri M., Thanh T.M.: Fundamental Incomplete Cryptography Method to Digital Rights Management Based on JPEG Lossy Compression. In: *2012 IEEE 26th International Conference on Advanced Information Networking and Applications*, pp. 755–762, 2012. doi: 10.1109/AINA.2012.111.

[7] Iwakiri M., Thanh T.M.: Incomplete Cryptography Method Using Invariant Huffman Code Length to Digital Rights Management. In: *2012 IEEE 26th International Conference on Advanced Information Networking and Applications*, pp. 763–770, 2012. doi: 10.1109/AINA.2012.112.

[8] Mary S.J.J., Joe S.S.A., Siddique M.: IRDM watermarking with EMRC6 encryption for DRM system. In: *2018 Majan International Conference (MIC)*, pp. 1–5, 2018. doi: 10.1109/MINTC.2018.8363157.

[9] Nakamoto S.: Bitcoin: A Peer-to-Peer Electronic Cash System. pp. 1–9, 2008. Available at https://bitcoin.org/bitcoin.pdf.

[10] Parah S.A., Sheikh J.A., Loan N.A., Bhat G.M.: Robust and blind watermarking technique in DCT domain using inter-block coefficient differencing, *Digital Signal Processing*, vol. 53, pp. 11–24, 2016. doi: 10.1016/j.dsp.2016.02.005.

[11] Pardhu T., Perli B.R.: Digital image watermarking in frequency domain. In: *2016 International Conference on Communication and Signal Processing (ICCSP)*, pp. 0208–0211, 2016. doi: 10.1109/ICCSP.2016.7754123.

[12] Ranjbar Alvar S., Akbari M., Yue D.M.X., Zhang Y.: NFT-Based Data Marketplace with Digital Watermarking. In: *KDD '23: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4756–4767, Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3580305.3599876.

[13] Srivastava R., Tomar R., Gupta M., Yadav A.K., Park J.: Image Watermarking Approach Using a Hybrid Domain Based on Performance Parameter Analysis, *Information*, vol. 12(8), 2021. doi: 10.3390/info12080310.

[14] Sultan K., Ruhi U., Lakhani R.: Conceptualizing Blockchains: Characteristics & Applications, *ArXiv*, vol. abs/1806.03693, 2018.

[15] Tai L.D., Thanh N.V., Thanh T.M.: Blockmarking: Hybrid Model of Blockchain and Watermarking Technique for Copyright Protection. In: *SoIC '22: Proceedings of the 11th International Symposium on Information and Communication Technology*, pp. 398–404, Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3568562.3568575.

[16] Tai L.D., Thanh T.M.: Digital Image Watermarking Algorithm Using Blockmarking Technique for Copyright Protection. In: *2023 15th International Conference on Knowledge and Systems Engineering (KSE)*, pp. 1–4, 2023. doi: 10.1109/KSE59128.2023.10299411.

[17] Thanh T.M., Iwakiri M.: A Proposal of Digital Rights Management Based on Incomplete Cryptography Using Invariant Huffman Code Length Feature, *Multimedia Systems*, vol. 20(2), pp. 127–142, 2014. doi: 10.1007/s00530-013-0327-z.

[18] Thanh T.M., Iwakiri M.: Fragile Watermarking with Permutation Code for Content-Leakage in Digital Rights Management System, *Multimedia Systems*, vol. 22, pp. 603–615, 2016. doi: 10.1007/s00530-015-0472-7.

[19] Thanh T.M., Quyet D.T.: A study on gas cost of ethereum smart contracts and performance of blockchain on simulation tool. In: *Peer-to-Peer Networking and Applications*, pp. 200–212, Special Issue on 2 – Track on Security and Privacy, 2024. doi: 10.1007/s12083-023-01598-3.

[20] Thanh T.M., Tanaka K.: A proposal of novel q-DWT for blind and robust image watermarking. In: *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, pp. 2061–2065, 2014. doi: 10.1109/PIMRC.2014.7136511.

[21] Thanh T.M., Tanaka K.: Blind Watermarking using QIM and the Quantized SVD Domain based on the *q*-Logarithm Function. In: J. Braz, S. Battiato, F.H. Imai (eds.), *VISAPP 2015 – Proceedings of the 10th International Conference on Computer Vision Theory and Applications, Volume 3, Berlin, Germany, 11–14 March, 2015*, pp. 14–25, SciTePress, 2015. doi: 10.5220/0005291900140025.

[22] Thanh T.M., Tanaka K.: The novel and robust watermarking method based on q-logarithm frequency domain, *Multimedia Tools and Applications*, vol. 75, pp. 11097–11125, 2016. doi: 10.1007/s11042-015-2836-6.

[23] Thanh T.M., Thanh N.T.: Extended DCT Domain for Improving the Quality of Watermarked Image. In: *2015 Seventh International Conference on Knowledge and Systems Engineering (KSE)*, pp. 336–339, 2015. doi: 10.1109/KSE.2015.70.

[24] Thomas T., Emmanuel S., Subramanyam A.V., Kankanhalli M.S.: Joint Watermarking Scheme for Multiparty Multilevel DRM Architecture, *IEEE Transactions on Information Forensics and Security*, vol. 4(4), pp. 758–767, 2009. doi: 10.1109/TIFS.2009.2033229.

[25] Venugopalan S., Aydt H.: Improving Confidentiality for NFT Referenced Data Stores, 2023.

[26] Zhaofeng M., Weihua H., Hongmin G.: A new blockchain-based trusted DRM scheme for built-in content protection, *EURASIP Journal on Image and Video Processing*, 91, 2018. doi: 10.1186/s13640-018-0327-1.

Affiliations

Le Danh Tai
Le Quy Don Technical University, 239 Hoang Quoc Viet, Bac Tu Liem, Ha Noi, ledanhtai@lqdtu.edu.vn

Ta Minh Thanh
Corresponding author. Le Quy Don Technical University, 239 Hoang Quoc Viet, Bac Tu Liem, Ha Noi, thanhtm@lqdtu.edu.vn

Received: 24.04.2023
Revised: 24.02.2024
Accepted: 9.06.2024

Computer Science • 25(4) 2024 https://doi.org/10.7494/csci.2024.25.4.6164

M Shanmuga Priya
Pavithra A
Leema Nelson

CHARACTER/WORD MODELLING: A TWO-STEP FRAMEWORK FOR TEXT RECOGNITION IN NATURAL SCENE IMAGES

Abstract *Text recognition from images is a complex task in computer vision. Traditional text recognition methods typically rely on Optical Character Recognition (OCR); however, their limitations in image processing can lead to unreliable results. However, recent advancements in deep-learning models have provided an effective alternative for recognizing and classifying text in images. This study proposes a deep-learning-based text recognition system for natural scene images that incorporates character/word modeling, a two-step procedure involving the recognition of characters and words. In the first step, Convolutional Neural Networks (CNN) are used to differentiate individual characters from image frames. In the second step, the Viterbi search algorithm employs lexicon-based word recognition to determine the optimal sequence of recognized characters, thereby enabling accurate word identification in natural scene images. The system is tested using the ICDAR 2003 and ICDAR 2013 datasets from the Kaggle repository, and achieved accuracies of 78.5% and 80.5%, respectively.*

Keywords scene text recognition, convolution neural network, character recognition, character/word modelling

Citation Computer Science 25(4) 2024: 637–652

1. Introduction

Text recognition in natural scene images is a critical task in the fields of computer vision and machine learning, which aims to build computer software that automatically extracts text from natural scene images. This technology has widespread applications in areas such as automated identification of traffic signals, license plates, and autonomous robot navigation [10, 22]. While numerous studies have focused on text recognition, most have focused on documents or digital paper-based materials, neglecting the complexities of extracting text from natural scene images. The intricate nature of this task arises from the diverse layouts and styles of characters, encompassing factors such as font, shape, size, color, and position while contending with challenges such as noise, blur, occlusions, and non-uniform lighting [25]. Earlier studies have used, OCR engines, such as ABBYY and TESSERACT, for text recognition, but their efficacy in scene image processing has been limited.

The main aim of this study is to recognize text present in printed images [12,18]. Many researchers have acknowledged the effectiveness of deep learning architectures for text recognition tasks [2, 6, 14]. These architectures contain multiple layers for various purposes including input representation, feature extraction, and classification. Among these, convolutional neural networks (CNN) have attracted significant attention for computer vision applications [10, 16, 19, 25]. The CNN architecture has several layers, including input, middle, and output layers. The input layer processes the inputs, whereas the middle layers with convolution and pooling features extract relevant information. Finally, an output layer with one or more fully connected layers performs the final classification [8].

This study conducted a comprehensive comparative analysis of text recognition methods that employ various deep learning architectures. To evaluate the effectiveness of the proposed system, two widely used public datasets are employed: the ICDAR 2003 and ICDAR 2013 datasets. The ICDAR 2013 dataset consisted of 229 training images and 233 testing images, each annotated at the word level. This dataset includes a diverse range of character and word graphics captured in natural settings and is suitable for various applications such as banners, displays, navigation panels, clothing, and house numbers. The selection of the sample images from the ICDAR 2013 dataset is shown in Figure 1.

The datasets selected for this task are characterized by a diverse range of images that showcase various sizes, scales, orientations, font types, and styles. These images are rich in characters and are carefully chosen to provide a comprehensive representation of subject matter. The proposed framework can be effectively evaluated in diverse authentic scenarios by incorporating various sample types, thereby enhancing its adaptability and credibility in recognizing text from natural scene images.

The main aim of this study is to achieve three objectives. The initial objective is to identify individual characters present in natural scene images using a CNN. The second objective involves recognizing the sequential order of text present in natural scene images, which is determined using a Viterbi search to determine the optimal

character sequence. The final objective is to conduct extensive experiments on two complex scene text recognitions using benchmark datasets to demonstrate the performance of the proposed system.

Figure 1. Natural scene images taken from the ICDAR 2013 dataset

The problem addressed by the proposed system is scene text recognition, which involves recognition of text in natural scene images captured by cameras or other devices. This is a challenging task because of various factors such as varying lighting conditions, complex backgrounds, and different fonts and languages [8, 19, 24]. Traditional methods for text recognition rely on handcrafted features and complex modelling, which are time consuming and may not be able to handle variability in real-world scenes. Therefore, there is a need for an automated system that can accurately recognize text in scene images, even in challenging scenarios.

The remainder of this paper is organized as follows. Section 2 presents prior research efforts and contextualizes the advanced scene-text recognition techniques. Section 3 details the design and implementation of the proposed system, delving into its architecture, algorithms, and key functionalities, while also demonstrating the process of creating a blueprint for scene-text recognition using character/word modeling. Section 4 evaluates and discusses the experimental results and findings of the proposed system by assessing both the character and word recognition modules. Finally, Section 5 concludes the implementation of character/word modeling as a means of text recognition in natural images.

2. Related works

This section provides an overview of the most advanced methods for scene-text recognition. The two common processes in text recognition are character-based and word-based processes. Character-based recognition depends on the detection and recognition of each character to identify an entire word when the characters are combined. Character segmentation and recognition are the bases of traditional character-based

recognition systems [16]. Several studies have employed a sliding window strategy that incorporates various scales for character segmentation and recognition. The recognition of scene text is challenging due to the intricacy of the task and the impact that segmentation, an essential component, has on the overall recognition system. The absence of segmentation in other character-based techniques for scene-text recognition highlights the difficulties of this process.

Recent methods, such as connectionist temporal classification, create character predictions followed by a post-processing stage [10]. Post-processing techniques that incorporate linguistic knowledge can be employed to enhance the accuracy of scene-text recognition. For example, Thillou et al. (2005) utilized n-gram scores to restrict an inference algorithm's prediction of a correct word [19]. Shi et al. (2016) proposed a deep architecture that integrates a convolutional neural network (CNN) with a recurrent neural network (RNN) to identify scene texts in images. Convolutional layers, which collect characteristics from the input picture, and recurrent layers, which predict a label, make up this design distribution for each frame, and a transcription layer based on connectionist temporal classification converts the frame predictions into a label sequence [17].

The main objective of a word-based recognition system is to obtain features from a complete word picture, without implementing character segmentation. The second objective is to integrate or pool these features into a predetermined architecture to conduct word classification and subsequently recognition [5]. Chen et al. (2020) developed an adaptive embedding gate for attention-based scene-text recognition to detect text based on neurocomputing. Fisher vectors are combined with pyramidal histograms of characters [1]. To develop a word-based recognition system, these vectors are combined with densely extracted low-level descriptors and spatial pyramids [9, 24]. Word recognition was achieved using the maximum posterior estimate obtained from a finite-state-weighted transducer. To address scene-text recognition, a 90k-class-based CNN was designed, where each class corresponds to a word in the lexicon [5].

In general, there are two basic strategies for using image-specific lexicons. A dictionary or lexicon of terms is included to enhance the effectiveness of the word-based recognition system. This list of candidate words allows the system to fix some of its error [4]. To ensure accuracy, it is recommended to use a powerful algorithm that searches for the dictionary term closest to the anticipated word. This approach is particularly important for word-based recognition, because it simultaneously captures both low-level features and high-level linguistic priors. A character recognition module can also be implemented in conjunction with a word recognition module to improve the results.

3. System design

This system design demonstrates the process of creating a blueprint for the development of scene-text recognition using character/word modelling. In this study, a deep

learning-based system is developed to recognize text embedded in natural-scene images. Figure 2 shows the proposed framework, which consists of three modules: image preprocessing, character recognition, and word recognition. In the first module, the input image is preprocessed to enhance the text recognition. In the second module, characters with different variations are recognized from the images using a popular CNN-based architecture. Finally, in the third module, Viterbi search algorithms are used to determine the best character sequence, which guides the system in determining the exact word in the image.

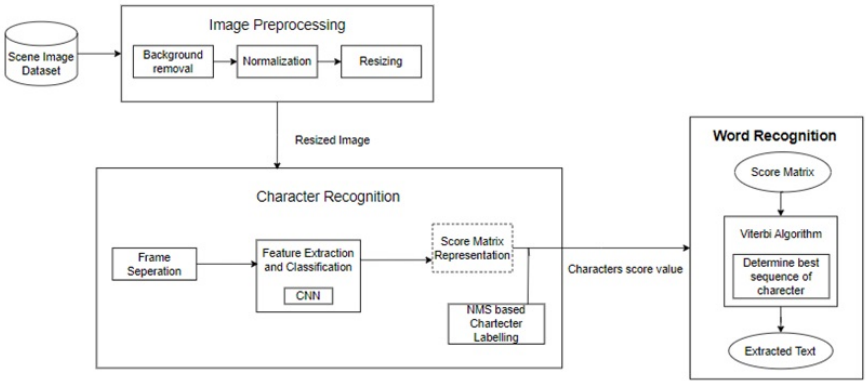


Figure 2. Overall system architecture

3.1. Image preprocessing module

The system accepts an image with text as input. The input image is preprocessed to enhance the quality and facilitate the extraction of text regions. Preprocessing refers to activities that involve pictures at the most fundamental level of abstraction. The input and output of the system are intensity images that are essentially identical to the original sensor data. Intensity images are typically depicted as matrices of image function values or brightness levels. The aim of preprocessing is to improve the picture data by removing unwanted distortions or enhancing the useful visual characteristics for further processing. Figure 3 shows the steps involved in image preprocessing, where the inputs are natural-scene images.

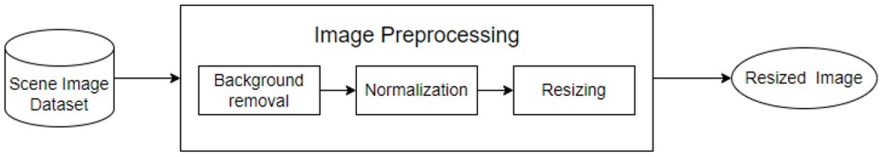


Figure 3. Image pre-processing

The steps involved are background removal, normalization, and image resizing. The image background removal step eliminates or alters the background from natural

scene images. Image normalization involves scaling the pixel values of an image to a fixed range or the mean and standard deviation, which can help mitigate the effects of lighting variations and other distortions. In the image-resizing step, each image is resized to a fixed height and width using a downsampling operation. This is because images captured from real-world scenes have character images of various scales, sizes, locations, and orientations. Therefore, to recognize characters in these images, it is mandatory to resize the image to a fixed height and width.

3.2. Character recognition module

In this module, the resized image is used as an input for character recognition. The image is divided into individual frames, which are then processed to identify the character region. The frames are separated from the resized image, and each frame is processed to localize the character region. To extract the important frame features, they must be processed using a cascaded CNN architecture. The CNN architecture consists of four stages of convolutional and subsampling layers used for feature extraction, whereas a fully connected layer is used for classification. The CNN architecture combines convolutional and subsampling layers to extract characteristics and predict class labels from input frames, motivated by the visual cortex structure. The filters detect specific characteristics in the input frame, thereby producing activation maps that are passed to the next layer in the CNN architecture [11]. Ultimately, the CNN outputs two crucial elements: the class label corresponding to the input frame and a score matrix representing the probabilities of the characters in each frame (p(char–I)). With characters falling into 62 possibilities (including 26 lowercase letters, 26 uppercase letters, and 10 numbers), it is important to note that a frame may encompass either the entire character or only part of it. Additionally, a character may span one or more frames at times. To address this challenge, a non-maximum-suppression method is applied. This method determines the frame containing the complete character based on the score value, thereby allowing the removal of redundant frames from classification inputs. The character-recognition process is illustrated in Figure 4.

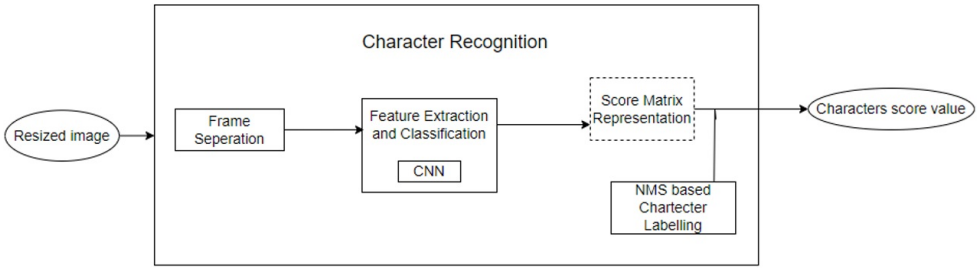


Figure 4. Character recognition

The score matrix depicted in Figure 5 represents the probabilities of the letters in the input image, which contains the word "NEW." The Viterbi search algorithm, a topic covered in the following module, selects letters with highlighted probabilities as the optimal sequence of characters to determine a word.

	N	n	...	e	E	...	W	w
0								
1								
...								
A								
...								
E				0.156	1.029			
...								
M								
N	0.456	0.021						
O								
...								
W							1.01	0.985

Figure 5. Score matrix for the word NEW

3.3. Word recognition module

In this module, a lexicon-based word recognition method is proposed to identify word W from an input image based on character sequence S generated from the character recognition module. The objective is to determine the sequence of characters with the highest probability. The character recognition module provides a set of N probabilities corresponding to the character class labels for each examined frame (N corresponds to the total number of character classes). The Viterbi search algorithm is used to determine the probability of an optimal sequence of characters. This algorithm is implemented to convert predicted probabilities into words. When confronted with the challenge of a sequence of M overlapping frames, it is essential to determine the most likely path for the optimal character sequence. The Viterbi search algorithm is effective in handling such scenarios using a score matrix to identify the character sequence with the highest score, ultimately declaring it as a recognized word [3, 23].

Figure 6 depicts the comprehensive process of the word recognition system, showing the recognized text within a boundary box. This box may overlap with the original image used for testing. To ensure clarity, the recognized word is transformed into text format and displayed separately.

Figure 7 provides a sample output of the text recognition process, offering a tangible representation of the system's capability to identify and display recognized words accurately.

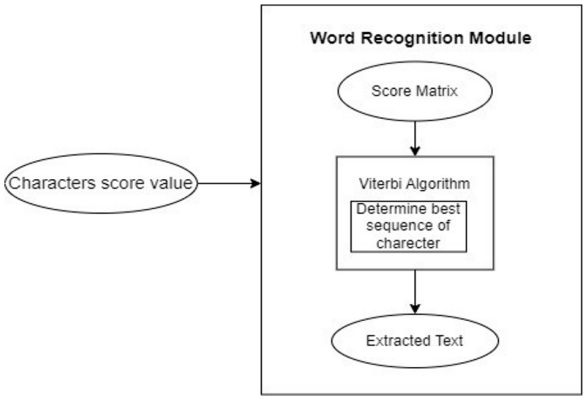


Figure 6. Word recognition

Figure 7. Output for word recognition

4. Results and discussion

This section discusses the experimental results and findings of the proposed system, by evaluating both the character and word recognition modules using the ICDAR 2003 and ICDAR 2013 datasets. The experiments are conducted on a computer with an Intel Core i7-7500U processor (2.9 GHz) and 8 GB RAM implemented using a Jupyter Source Notebook in Python. The performance assessment unfolds in two stages: character recognition, where various deep learning architectures, such as CNN, ANN, and RNN, are explored and compared, and word recognition, where the accuracy of the system is evaluated based on precision. The accuracy metric measures the ratio of correctly recognized characters or words to the total number of characters or words based on the ground truth.

Figure 8 illustrates the correct recognitions for each alphabetical character (A–Z, a–z) and number (0–9).

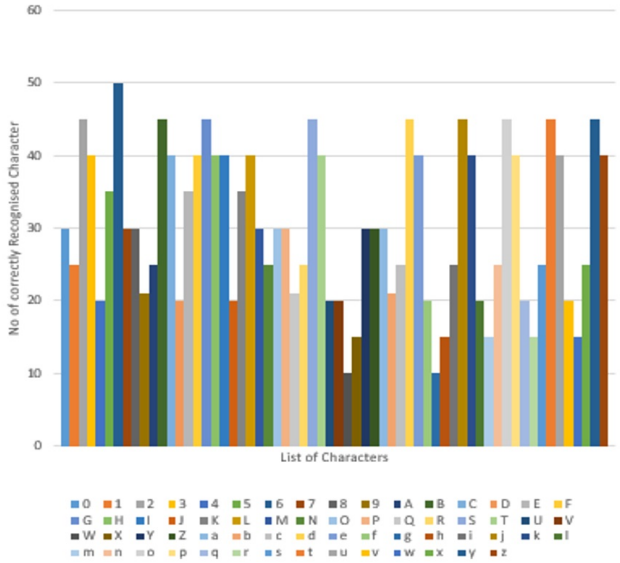


Figure 8. No. of times correctly recognised characters

The experimental results in Table 3 and Figure 9 highlight the superior performance of the proposed CNN-based text recognition method, which achieves the highest accuracy of 80.5% on the ICDAR 2003 dataset and 78.5% on the ICDAR 2013 dataset. Furthermore, the system performance is assessed by varying the width and height of the resized images, as listed in Tables 1 and 2.

Table 1
Word recognition accuracy in ICDAR 2003 different heights and widths

Height	Width	Accuracy [%]
80	80	79.8
90	90	80.7
100	100	81.5
150	150	77.2
200	200	76.3
300	200	71.5

Table 2
Word recognition accuracy in ICDAR 2013 with different heights and widths

Height	Width	Accuracy [%]
80	80	78
90	90	78.8
100	100	79.8
150	150	75.6
200	200	74.5
300	300	72.3

Table 3
Accuracy for test data using CNN

Datasets	Accuracies [%]
ICDAR 2003	80.5
ICDAR 2013	78.5

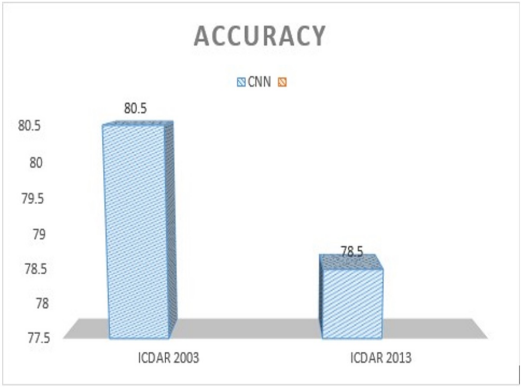


Figure 9. Accuracy for test data using CNN

Notably, the width and height of $(100, 100)$ demonstrates optimal accuracy, achieving 81.5% for the ICDAR 2003 dataset and 79.8% for the ICDAR 2013 dataset. Figure 10 and 11 show the word recognition accuracy for different heights and widths, revealing that the resized image of $(100, 100)$ outperforms other sizes.

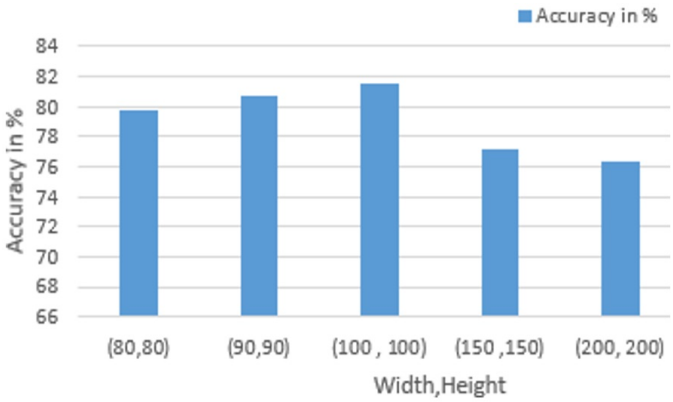


Figure 10. Word recognition accuracy in ICDAR 2003 with different heights and widths

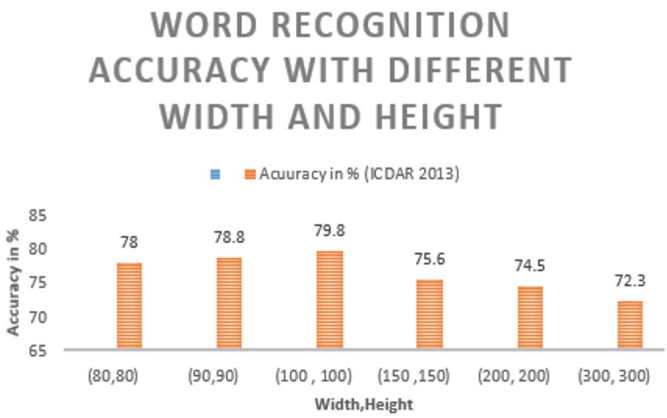


Figure 11. Word recognition accuracy in ICDAR 2013 with different heights and widths

The robustness of the system is tested against various image qualities, such as clear, blurred, and low-quality images, as shown in Table 4 and Figure 12.

Table 4
Precision value for different categories of Image data

Image category	ICDAR 2003 [%]	ICDAR 2013 [%]
Low quality images	20	15
Blurred images	65	70
Cleared images	97	98

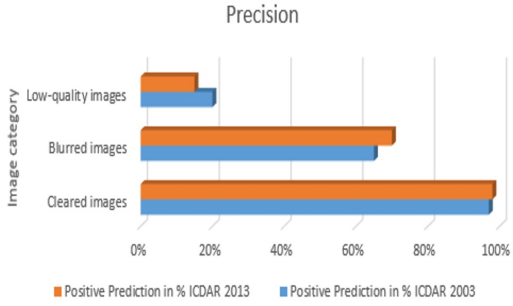


Figure 12. Precision value for different categories of image data

The system exhibited superior performance for clearer images. Comparisons between three different deep learning architectures CNN (Proposed Method), RNN, and ANN revealed that CNN outperformed the others. This is evident in Table 5 and Figure 13, where CNN achieved the highest accuracy in character recognition. Similarly, in word recognition, as shown in Figure 14 and Table 5, CNN outperformed the

other architectures. In summary, the proposed system, primarily utilizing a CNN, demonstrated remarkable accuracy in recognizing characters and words, showcasing resilience against varying image qualities, and confirming its efficacy in real-world scenarios.

Table 5
Word recognition accuracy

Datasets	CNN [%]	RNN [%]	ANN [%]
ICDAR 2003	80.5	78.9	70
ICDAR 2013	78.5	70.8	58.75

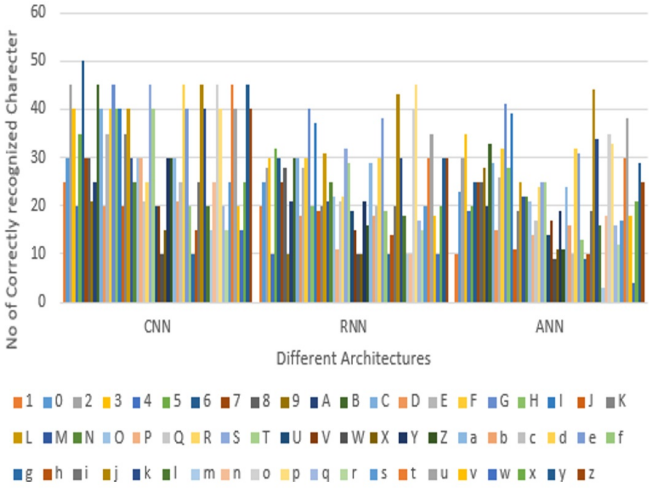


Figure 13. No. of correct character recognition using three different deep learning architectures

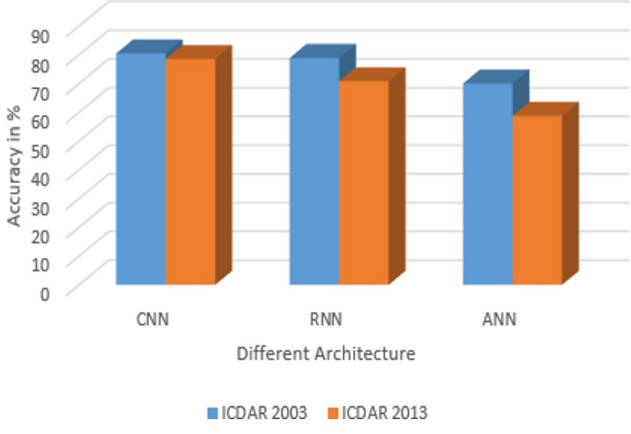


Figure 14. Word recognition accuracy

The result of the statistical test, as indicated by the obtained p-value of 0.235, demonstrated the level of significance associated with the contrast in accuracy between the two datasets under consideration. If the p-value is less than the predetermined significance level (e.g., 0.05), it can be reasonably inferred that there is a statistically significant difference in the performances of the two datasets. Statistical tests comparing the accuracies achieved using the ICDAR 2003 and ICDAR 2013 datasets are presented in Table 6.

Table 6
Statistical tests accuracy comparison using ICDAR 2003 and ICDAR 2013 datasets

S. No.	Dataset	Accuracy [%]	p-value
1	ICDAR 2003	80.5	0.235
2	ICDAR 2013	78.5	

The proposed system for scene text recognition demonstrated impressive accuracy and computational efficiency with notable speed, which makes it a strong candidate for practical applications. To provide a basis for comparison, the runtime of the system is assessed against those of state-of-the-art systems. The entire pipeline is implemented on an Intel Core i7 2.9 GHZ machine, yielding a runtime of approximately 0.30 seconds per image with GPU and 0.70 seconds per image with CPU. These results are competitive with other existing systems, such as the one reported in [4], which achieved an average runtime of 0.92 seconds per sample using a Core i5 2.80 GHZ processor. Another system, developed by Novikova et al. [15], had a speed of 0.85 seconds per image. In a separate study, [12] reported a runtime of approximately 0.25 seconds per image with GPU and 0.83 seconds per image with CPU only on a computer with an Intel Xeon E5 2.6 GHZ x 2 processor. The proposed word recognition system is compared with existing state-of-the-art techniques, as shown in Table 7.

Table 7
Comparison between proposed word recognition system with existing state of art systems

S. No.	Methods	ICDAR 2003 dataset accuracy [%]	ICDAR 2013 dataset accuracy [%]
1	[21]	62.00	70.00
2	[7]	66.19	–
3	[15]	–	72.40
4	[20]	80.56	–
5	[13]	78.20	–
6	[4]	78.44	82.31
7	[9]	81.70	79.40
8	Proposed system	80.50	78.50

5. Conclusion

This study implemented character/word modelling to recognize texts in natural images. This two-step process demonstrates significant advancements in scene-text recognition. In the first step, the CNN model is used for character recognition, and the Viterbi search algorithm is used for word recognition in the second step. The developed model is tested using ICDAR 2003 and ICDAR 2013 datasets. The results demonstrated the effectiveness of the cascaded CNN architecture for recognizing characters and words in various scenarios. Moreover, the developed model adapts to different image qualities, such as clear, blurred, and low-quality images, thereby highlighting its robustness in real-world applications. The results obtained from the comparative analysis show that existing deep learning techniques, such as RNN, ANN, and CNN, achieve the highest accuracy for character recognition. The developed model not only surpasses traditional optical character recognition (OCR) techniques but also shows promising results in challenging scenarios, making it suitable for applications such as document digitization, number plate recognition, image-based search engines, and augmented reality.

By exploring diverse convolutional neural network (CNN) architectures, including transformer-based models and innovative convolutional layers, it may be possible to enhance the performance in character and word recognition through improved feature extraction and context comprehension. This can be achieved by incorporating attention mechanisms and advanced techniques.

References

[1] Almazán J., Gordo A., Fornés A., Valveny E.: Word Spotting and Recognition with Embedded Attributes, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36(12), pp. 2552–2566, 2014. doi: 10.1109/tpami.2014.2339814.

[2] Arafat S.Y., Iqbal M.J.: Urdu-Text Detection and Recognition in Natural Scene Images Using Deep Learning, *IEEE Access*, vol. 8, pp. 96787–96803, 2020. doi: 10.1109/access.2020.2994214.

[3] Bahi H.E., Zatni A.: Text recognition in document images obtained by a smartphone based on deep convolutional and recurrent neural network, *Multimedia Tools and Applications*, vol. 78(18), pp. 26453–26481, 2019. doi: 10.1007/s11042-019-07855-z.

[4] Bhunia A.K., Kumar G., Roy P.P., Balasubramanian R., Pal U.: Text recognition in scene image and video frame using color channel selection, *Multimedia Tools and Applications*, vol. 77, pp. 8551–8578, 2018. doi: 10.1007/s11042-017-4750-6.

[5] Chen X., Wang T., Zhu Y., Jin L., Luo C.: Adaptive embedding gate for attention-based scene text recognition, *Neurocomputing*, vol. 381, pp. 261–271, 2020. doi: 10.1016/j.neucom.2019.11.049.

[6] Coates A., Carpenter B., Case C., Satheesh S., Suresh B., Wang T., Wu D.J., Ng A.Y.: Text detection and character recognition in scene images with unsupervised feature learning. In: *2011 International Conference on Document Analysis and Recognition*, pp. 440–445, IEEE, 2011. doi: 10.1109/icdar.2011.95.

[7] Elagouni K., Garcia C., Mamalet F., Sebillot P.: Combining multi-scale character recognition and linguistic knowledge for natural scene text OCR. In: *DAS '12: Proceedings of the 2012 10th IAPR International Workshop on Document Analysis Systems*, pp. 120–124, IEEE, 2012. doi: 10.1109/das.2012.26.

[8] Goel V., Mishra A., Alahari K., Jawahar C.V.: Whole is greater than sum of parts: Recognizing scene text words. In: *2013 12th International Conference on Document Analysis and Recognition*, pp. 398–402, IEEE, 2013. doi: 10.1109/icdar.2013.87.

[9] Harizi R., Walha R., Drira F., Zaied M.: Convolutional neural network with joint stepwise character/word modelling based system for scene text recognition, *Multimedia Tools and Applications*, pp. 3091–3106, 2022. doi: 10.1007/s11042-021-10663-z.

[10] Jaderberg M., Simonyan K., Vedaldi A., Zisserman A.: Reading text in the wild with convolutional neural networks, *International Journal of Computer Vision*, vol. 116, pp. 1–20, 2016. doi: 10.1007/s11263-015-0823-z.

[11] Liao M., Zhang J., Wan Z., Xie F., Liang J., Lyu P., Yao C., Bai X.: Scene text recognition from two-dimensional perspective. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 8714–8721, 2019. doi: 10.1609/aaai.v33i01.33018714.

[12] Liu X., Kawanishi T., Wu X., Kashino K.: Scene text recognition with CNN classifier and WFST-based word labeling. In: *2016 23rd International Conference on Pattern Recognition (ICPR)*, pp. 3999–4004, IEEE, 2016. doi: 10.1109/icpr.2016.7900259.

[13] Liu X., Kawanishi T., Wu X., Kashino K.: Scene text recognition with high performance CNN classifier and efficient word inference. In: *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1322–1326, IEEE, 2016. doi: 10.1109/icassp.2016.7471891.

[14] Long S., He X., Yao C.: Scene Text Detection and Recognition: The Deep Learning Era, 2018, arXiv preprint arXiv:181104256. arXiv:1811.04256.

[15] Novikova T., Barinova O., Kohli P., Lempitsky V.: Large-lexicon attribute-consistent text recognition in natural images, *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision Florence, Italy, October 7–13, 2012 Proceedings, Part VI*, pp. 752–765, 2012. doi: 10.1007/978-3-642-33783-3_54.

[16] Portaz M., Kohl M., Chevallet J.P., Quénot G., Mulhem P.: Object instance identification with fully convolutional networks, *Multimedia Tools and Applications*, vol. 78(3), pp. 2747–2764, 2019. doi: 10.1007/s11042-018-5798-7.

[17] Shi B., Bai X., Yao C.: An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39(11), pp. 2298–2304, 2016. doi: 10.1109/tpami.2016.2646371.

[18] Shivakumara P., Bhowmick S., Su B., Tan C.L., Pal U.: A new gradient-based character segmentation method for video text recognition. In: *2011 International conference on document analysis and recognition*, pp. 126–130, IEEE, 2011. doi: 10.1109/icdar.2011.34.

[19] Thillou C., Ferreira S., Gosselin B.: An embedded application for degraded text recognition, *EURASIP Journal on Advances in Signal Processing*, 370317, 2005. doi: 10.1155/asp.2005.2127.

[20] Wang D.H., Wang H., Zhang D., Li J., Zhang D.: Robust scene text recognition using sparse coding based features, 2015. ArXiv preprint arXiv:1512.08669., arXiv:1512.08669.

[21] Wang K., Babenko B., Belongie S.: End-to-end scene text recognition. In: *2011 International Conference on Computer Vision*, pp. 1457–1464, IEEE, 2011-11. doi: 10.1109/iccv.2011.6126402.

[22] Xu C., Yang J., Gao J.: Coupled-learning convolutional neural networks for object recognition, *Multimedia Tools and Applications*, vol. 78, pp. 573–589, 2019. doi: 10.1007/s11042-017-5262-0.

[23] Xue C., Huang J., Zhang W., Lu S., Wang C., Bai S.: Image-to-character-to-word transformers for accurate scene text recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45(11), pp. 12908–12921, 2023. doi: 10.1109/tpami.2022.3230962.

[24] Yuan J., Wei B., Liu Y., Zhang Y., Wang L.: A method for text line detection in natural images, *Multimedia Tools and Applications*, vol. 74, pp. 859–884, 2015. doi: 10.1007/s11042-013-1702-7.

[25] Zhang Z., Zhang C., Shen W., Yao C., Liu W., Bai X.: Multi-oriented Text Detection with Fully Convolutional Networks. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4159–4167, 2016. doi: 10.1109/cvpr.2016.451.

Affiliations

M Shanmuga Priya
Anna University, Chennai, Tamil Nadu, India, mscsepriya@annauniv.edu

Pavithra A
Anna University, Chennai, Tamil Nadu, India, spriya@cs.annauniv.edu

Leema Nelson
Chitkara University, Chitkara University Institute of Engineering & Technology, India, Punjab, leema.nelson@gmail.com

Received: 12.03.2024
Revised: 11.05.2024
Accepted: 5.06.2024

Information for Authors

We accept only the original scientific papers prepared in English. The papers are to be prepared using the LaTeX system. Submitted papers will be refereed by independent reviewers and, if necessary, the Authors may be asked to revise their manuscripts. Proofs will be sent to the Authors for corrections. There is no publication fee. Authors of the accepted papers are eligible to get one hardcopy of the volume containing their contribution free of charge. No postage charges apply.

Our website

https://journals.agh.edu.pl/csci/

Open access

This is an open access journal which means that all content is freely available without charge to the user or his/her institution. This is in accordance with the Budapest Open Access Initiative definition of open access. All printed volumes may be accessed at our website.

Indexing

Google Scholar
http://scholar.google.com

Index Copernicus
http://indexcopernicus.com/

Directory of Open Access Journals
http://www.doaj.org

Open Archives Initiative
http://www.openarchives.org

Digital Libraries Federation
http://fbc.pionier.net.pl/owoc/

BazTech
http://baztech.icm.edu.pl

Worldcat
http://www.worldcat.org

WorldWideScience.org
http://worldwidescience.org

Sherpa Romeo
http://www.sherpa.ac.uk/romeo/

Journal TOCs
http://www.journaltocs.ac.uk

SCOPUS
www.scopus.com

Web of Science – Emerging Sources Citation Index
www.webofknowledge.com

The Computer Science Journal is published by the AGH University of Science and Technology, Krakow Poland. The Editors of the Journal are members of the Faculty of Computer Science, Electronics and Telecommunications and the Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering. The Editorial Board consists of many renowned computer science researchers from all over the world.

The first issue of the Journal was published in 1999. Currently, the Journal is published quarterly, with the main goal to create a forum for exchanging research experience for scientists specialized in different fields of computer science.

Original papers are sought concerning theoretical and applied computer science problems. Example areas of interest are:

- theoretical aspects of computer science,
- pattern recognition and processing,
- evolutionary algorithms,
- neural networks,
- database systems,
- knowledge engineering,
- automatic reasoning,
- computer networks management,
- distributed and grid systems,
- multi-agent systems,
- multimedia systems and computer graphics,
- natural language processing,
- soft-computing,
- embedded systems,
- adaptive algorithms,
- simulation.

Abstracts, full versions of the issued volumes and instructions for authors and reviewers may be found at http://csci.agh.edu.pl