

Article

A New COVID-19 Detection Method Based on CSK/QAM Visible Light Communication and Machine Learning

Ismael Soto ^{1,*}, Raul Zamorano-Illanes ¹, Raimundo Becerra ², Pablo Palacios Játiva ^{2,3}, Cesar A. Azurdia-Meza ², Wilson Alavia ¹, Verónica García ⁴, Muhammad Ijaz ⁵ and David Zabala-Blanco ⁶

¹ CIMTT, Department of Electrical Engineering, Universidad de Santiago de Chile, Santiago 9170124, Chile

² Department of Electrical Engineering, Universidad de Chile, Santiago 8370451, Chile

³ Escuela de Informática y Telecomunicaciones, Universidad Diego Portales, Santiago 8370190, Chile

⁴ Departamento en Ciencia y Tecnología de los Alimentos, de la Universidad de Santiago de Chile, Santiago 9170124, Chile

⁵ Manchester Metropolitan University, Manchester M1 5GD, UK

⁶ Department of Computer Science and Industry, Universidad Católica del Maule, Talca 3480112, Chile

* Correspondence: ismael.soto@usach.cl

Abstract: This article proposes a novel method for detecting coronavirus disease 2019 (COVID-19) in an underground channel using visible light communication (VLC) and machine learning (ML). We present mathematical models of COVID-19 Deoxyribose Nucleic Acid (DNA) gene transfer in regular square constellations using a CSK/QAM-based VLC system. ML algorithms are used to classify the bands present in each electrophoresis sample according to whether the band corresponds to a positive, negative, or ladder sample during the search for the optimal model. Complexity studies reveal that the square constellation $N = 2^{2^i} \times 2^{2^i}$, ($i = 3$) yields a greater profit. Performance studies indicate that, for BER = 10^{-3} , there are gains of -10 [dB], -3 [dB], 3 [dB], and 5 [dB] for $N = 2^{2^i} \times 2^{2^i}$, ($i = 0, 1, 2, 3$), respectively. Based on a total of 630 COVID-19 samples, the best model is shown to be XGBoots, which demonstrated an accuracy of 96.03%, greater than that of the other models, and a *recall* of 99% for positive values.

Keywords: COVID-19; CSK; QAM; VLC; BER



Citation: Soto, I.; Zamorano-Illanes, R.; Becerra, R.; Palacios Játiva, P.; Azurdia-Meza, C.A.; Alavia, W.; García, V.; Ijaz, M.; Zabala-Blanco, D. A New COVID-19 Detection Method Based on CSK/QAM Visible Light Communication and Machine Learning. *Sensors* **2023**, *23*, 1533. <https://doi.org/10.3390/s23031533>

Academic Editor:
Christian-Alexander Bunge

Received: 24 October 2022
Revised: 23 January 2023
Accepted: 25 January 2023
Published: 30 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

While the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is still undergoing new mutations, it is currently risky to declare that the virus is no longer a problem. It is unknown whether the current vaccines prevent severe symptoms, hospitalisation or death. Some research on new COVID-19 variants indicates that the virus is spreading faster than in the past [1] and has effects of maternal SARS-CoV-2 infection on pregnant women, fetuses, and newborns [2]. A study on the presence of COVID-19 and its association with respiratory syncytial virus was conducted during the winter of 2020–2021 in Europe and North America [3] to determine whether SARS-CoV-2 mutates similarly globally or whether it mutates differently in specific populations [4]. Newly emerging variants of SARS-CoV-2 continue to pose a significant threat to global public health by causing COVID-19 epidemics [5]. The SARS-CoV-2 pandemic has highlighted the need for routine monitoring of infections in high-density indoor areas, such as hospitals and underground environments, with the strictest monitoring required for dust particles in pollution-absorbing tunnels and metro stations.

The primary method for preventing transmission is social distancing, for which measurement mechanisms have been developed that establish areas of risk based on the number of people in a given geographical area [6]; however, the collection of medical samples does not allow the health system to maintain constant monitoring, because it requires patients to

visit a medical centre or have medical personnel visit their home. In addition, only variables such as temperature [7] and physical conditions can be monitored continuously [8].

This strategy was tested in many areas during the SARS-CoV-2 pandemic and was validated by a number of authors [9,10]. Another benefit of pathogen identification through sewage is the ability to monitor both endemic and Waterborne Datasets (WBDs) [9,11–13]. The latter are collections of microorganisms primarily related to diarrhoea disorders that are transferred by water or food irrigated in polluted water and generate large-scale outbreaks. A national pathogen monitoring system utilising Optical Wireless Communications (OWC) technologies will provide a significant contribution and is an innovative technique for detecting these pathogens. Individual testing and traceability systems are one way to achieve this objective. However, the cost of the other detection methods and the limited representativeness of the gathered data prevent the creation of appropriate models for this objective. It reduces the number of samples required for analysis, is representative of the population whose waste is channelled into the sample, is independent of sanitary conditions due to the low availability of tests, enables the observation of under-represented or asymptomatic diseases, and is less expensive than other methods [14]. The use of mathematical methods on images in conjunction with ML techniques yields results that aid in less subjective decisions being made, allowing later validation of the diagnosis [15]. Among other things, the impacts of the epidemic on public health, culture, the environment, and the economy [16] have provided motivation for the use of technologies such as Artificial Intelligence (AI), ML [17], robotics [18], big data [19], and the IoT [16]. In order to measure vital signs, many of the mechanisms employed for this purpose include equipment that must come into contact with the body of an individual [20,21] by using specimens from medical facilities [22,23] or at home through the use of robots [18].

In relation to this paper, our team has published original models of underground channels [24–28]. Additionally, before the pandemic, we produced papers on DNA [29–31] and work on MIMO [32–34]. Pathogens will always be around humans, so it is crucial to conduct channel studies that allow us to transfer information under any circumstance in a secure and fast manner.

This manuscript makes the following contributions to this area of research: It provides a mathematical model for mapping DNA genes using a CSK/QAM scheme transmitted by Frequency Shift Keying (FSK) over a MIMO VLC-based underground channel and an ML-based technique for identifying COVID-19.

In addition to this introduction, there are four more sections in this document. In Section 2, the current state of knowledge is presented. Section 3 provides the methodology. Section 4 is a discussion of the results, while Section 5 contains the conclusion.

2. State-of-the-Art Techniques

This section presents the state-of-the-art techniques used for the application of VLC in underground channels. This is followed by the presentation of a model of a channel based on colour shift keying with quadrature Amplitude modulation (CSK/QAM) mapping, the use of the Galois Field Mapping/Galois Fields Demapping interface between the human side of the machine and the signal processing communication channel, and finally, AI-based procedures to replace human processing. Note that COVID-19 and pathogens in general are most likely to spread rapidly in indoor settings, such as hospitals or industrial settings like mine tunnels. As evidenced by events of the last three years, this can result in a decrease in a country's GDP and crippling of its economy.

2.1. Work Related to the VLC Channel

As mentioned in the previous paragraph, the fact that gene information can be colour-coded makes it necessary to discuss alternatives for the implementation of the underground channel. The above model can be used in interior settings, hospitals and underground tunnels. For this reason, studies on underground channels, scattering distribution patterns, and FSK in an underground channel are presented. Using experimental tests and math-

emational simulations, it has been found that FSK is the best method for reaching longer distances, because the energy is concentrated in a single frequency tone in a mining tunnel.

2.1.1. Work Related to Underground Channels

Earlier efforts modelled the VLC channel in underground mining environments using the same Lambertian channel model as an indoor VLC channel. This meant that dust scattering, reflections on uneven walls, light obstruction, also known as shadowing, and the relative tilt and rotation of LEDs and PDs were disregarded. Furthermore, light has a dual nature and, depending on the quantum interpretation of the observer, can be modelled as either a particle or a wave [35]. The majority of current research on VLC communication in underground channels focuses on on-off keying (OOK) modulation, which is similar to amplitude modulation. In terms of phase, frequency modulation includes both coherent and noncoherent signals. Coherent FSK signals are those whose phase remains constant over time. Inconsistency also exists when the phase changes or varies over time. The continuous use of pneumatic hammers to excavate rocks in underground tunnels generates a substantial quantity of airborne dust that is detrimental to any channel. Prior generation processes have always sought to include [36] VLC or hybrid communication. To model the scattering effect and incorporate it into the overall model of the underground mining visible light communication channel (UM-VLC), a robust mathematical infrastructure is required for UM-VLC [37,38]. Consequently, the effect of dust particles in the air is disregarded during testing in nonscattering indoor environments, such as offices, homes and hospitals; consequently, such models cannot function in underground tunnels.

2.1.2. Work on Scattering Distribution Patterns

When the dust particle size is very small, the proposed dispersion distribution models can be implemented in any type of multipath wireless communication system. However, as the dust particle size increases and the dust concentration causes small holes in the sensors, the waves revert to particle behaviour and scatter more. Previous studies modelled different arrival schemes over time and various communication scenarios [39–41]. A dust disc around the optical receiver was modelled as a uniform distribution within a 2D disc region [42,43]. In [40], a statistical analysis was conducted in a hemispheric area around a base station. Using a geometric model of a mobile transmitter channel, the signal's arrival time and direction were analysed. A Gaussian scatter distribution model was presented. Due to the model's spatial-temporal properties, in terms of the arrival angle and arrival time [41], it will eventually be possible to apply it to multipath wireless communication systems. Tennskoon [44] proposed a three-dimensional (3D) stochastic geometry model with a Gaussian distribution centred on an arbitrary point within a sphere.

2.1.3. Work Related to FSK in Underground Channels

As previously stated, the majority of the literature on VLC in general has used OOK and laboratory level tests at very short distances with white light. On the ground, this results in large, power-hungry drivers. Therefore, it might be interesting to include the frequencies or wavelengths of the chosen colours [45]. There are few FSK applications for VLC which, by definition, has a longer range and lower power consumption. In [46], an advertising panel that uses FSK to communicate with a cell phone application is discussed. Salmento [47] described a lab-scale VLC system comprised of a single-stage buck-boost power factor correction converter operating in discontinuous drive mode with dimming capability. Dahri [48] described a system for vehicle-to-vehicle communication using FSK. The only studies on FSK modulation applied to underground mining are presented in [49–51], which all involved testing in a mine tunnel.

The models discussed previously are extremely rigid and linearly conceptualised. Quantum objects, on the other hand, do not need to have their properties defined; a beam of light can arrive at the photodetector not only in a straight line in a coherent manner but

also via other angles in an incoherent manner [52]. Dust, for instance, can be modelled as large particles colliding with waves, which cause collisions that deflect the waves, but photon jets arrive at the photodetector because they are aligned. If these become stuck in the detector and clogged by dust, leaving a few holes that convert the photon jets back into waves, all of the previously discussed models become invalid. According to Aharonov [53], infrastructure maintenance and cleanliness appear to be more important than simple model validation.

2.1.4. Model of a Channel Based on CSK/QAM Mapping

In the context of visible light communication, the use of the colour shift keying modulation scheme has allowed the application of different techniques in relation to the optimisation of the colour space which, according to the IEEE 802.15.7 standard, presents nine valid schemes for the combination of colours with 4, 8 and 16 Color Shift Keying (CSK) constellations. However, performance should be improved by concatenating other modulation or coding methodologies. To improve spectrally efficient transmissions in VLC systems using CSK communication, coding and mapping techniques, such as bit-interleaved coded modulation with iterative demapping and decoding [54], are applied. These work best for high-speed VLC applications. Machine learning is also applied to find the most optimal combination of coding and modulation technologies. Its performance is based on the mapping of symbol permutations through points in an optimized CSK constellation, This offers benefits in terms of diversity, resistance to channel degradation monochrome, and increased security. This type of method can be combined with MIMO technologies [54] to evaluate systems through Monte Carlo simulations. For the generation of a multiuser channel, the optimisation of channel resources and the technical type of spatial division used are essential. Separation by means of CSK modulation techniques to maximise the minimum Euclidean distance between different points of a constellation or multiuser joint constellation [55] is used for this purpose. Techniques such as multiplexing of the symbols used in wireless channels where each 7-bit 128 QAM symbol is multiplexed by a complex value signal to form a 32 QAM with an additional gain of 40% is done to compensate for problems related to chromatic dispersion and non-Kerr linearity [56]. The use of constellation probability shaping is a high-order modulation format optimisation technology that optimises the probability distribution of each signal constellation point to improve the generalised mutual information and increase the transmission capacity of QAM modulation [57].

2.2. Galois Field Mapping/Galois Fields Demapping

In [58], the author describes the problems faced and efforts to eradicate the COVID-19 pandemic. In order to achieve this objective, documentation is produced to examine the signatures of genomes using chaotic studies. First, alternative representations of the SARS-COV-2 DNA sequences, such as colour-coded images, indicator matrices, DNA walks, and chaotic games were created.

In [59], the detection of cancer using images is proposed. Cells are constantly exposed to numerous mutagens that produce diverse types of DNA lesions. Eukaryotic cells have evolved to contain a vast array of DNA repair mechanisms that are capable of detecting and repairing these lesions, thereby preventing genomic instability. Based on their functions, repair proteins are recruited to lesions sequentially.

In [60], the helitrons, eukaryotic transposable elements transposed by the rolling-circle mechanism, are defined. These have been identified in numerous species with highly variable copy numbers and, in some cases, they comprise a significant portion of the genome. Using images of the constituent helitron features and a pretrained neural network as a classifier, classification was conducted using the k -means features corresponding to genomic sequences, and this method was compared with the Support Vector Machine (SVM) and Random Forest methods.

A few studies have employed Galois fields to numerically represent DNA [29,31]. Representation through Galois Fields $GF(p)$ is based on gel electrophoresis, a standard method for separating double-stranded DNA (dsDNA) fragments of different sizes previously obtained by the Polymerisation Chain Reaction. When interpreting the electrophoresis of $GF(p)$ and its extension $GF(p^n)$, the standard notation [61,62] is utilised. Pathogen-causing SARS-CoV-2 DNA is used for COVID-19 detection because it contains four distinct genes: Adenine (A), Cytosine (C), Guanine (G), and Thymine (T). By using four non-binary symbols, the four states $2^i, i = 1$ can be represented by natural numbers such as [0, 1, 2, 3] or by colours such as [red, green, blue, yellow] [63,64].

In [59], the author proposed the detection of cancer using images. Cells are constantly exposed to a variety of mutagens that generate various types of DNA lesions. In order to prevent genomic instability, eukaryotic cells have evolved a vast array of DNA repair mechanisms capable of detecting and repairing these lesions. According to their function, repair proteins are sequentially recruited to lesions.

Reference [60] described how helitrons, eukaryotic transposable elements (tes) transposed by the rolling-circle mechanism, have been identified in numerous species with highly variable copy numbers and, in some cases, constitute a significant portion of the genome. Using images of the constituent helitron features and a pretrained neural network as a classifier, classification using the k -means features that correspond to genomic sequences was conducted, and a comparison to the SVM and Random Forest methods was made.

Many studies have linked ML to the diagnosis of COVID-19 using lung X-rays [23,65,66]. Deep neural networks (DNN) were used to process images in [67], and statistical methods were used in conjunction with heuristic filtering to identify somatic mutations in tumour samples.

2.3. AI-Based Procedures to Replace Human Processing

In the manual detection of SARS-CoV-2, a machine–human couple interpreted the gel electrophoresis results following a two-step end point Reverse Transcriptase PCR (RT-PCR). In this method, the N_1 and N_2 gene targets are followed to detect SARS-CoV-2, and Ribonuclease P (*RNase P*) is used for Ribonucleic acid (RNA) extraction. A dataset of 242 gel images obtained in that study was utilised in this work [68].

Using a histogram database, [69] contributed to the formation of a stratification system with three severity levels (moderate, severe and mild) that defines infection in various slides from a COVID-19 patient. The authors of [70] argue that the use of deep learning in medical imaging is an emerging technology for the diagnosis of a variety of diseases, such as pneumonia, lung cancer, brain stroke, and breast cancer. Before constructing a predictive model, machine learning and conventional data mining techniques perform the time-consuming feature extraction process. A convolutional neural network (CNN) was constructed using 1920 Chest X-rays (CxR) from healthy individuals and COVID-19 infected patients as training data. Using the clinical results of the 300-CxR validation dataset, the performance of the developed CNN was assessed further.

3. System Diagram

Figure 1 is a system diagram that illustrates the phases of model searching and operation of the classification model. Assuming the diagram can be folded along the dashed line, the first three boxes at the emitter and the last three boxes at the receiver represent human activity that could be automated. The organic samples consist of chromosomes with four genes represented by four symbols, Adenine (A), Cytosine (C), Guanine (G), and Thymine (T), which are passed through thermocycles or undergo amplification of the deoxyribonucleic acid (DNA), separation and delivery to the next block. The "(+)", "(−)", and "Ladder" states of the receiver's reverse function, designated by the listing symbol, reveal the genes that contain COVID-19.

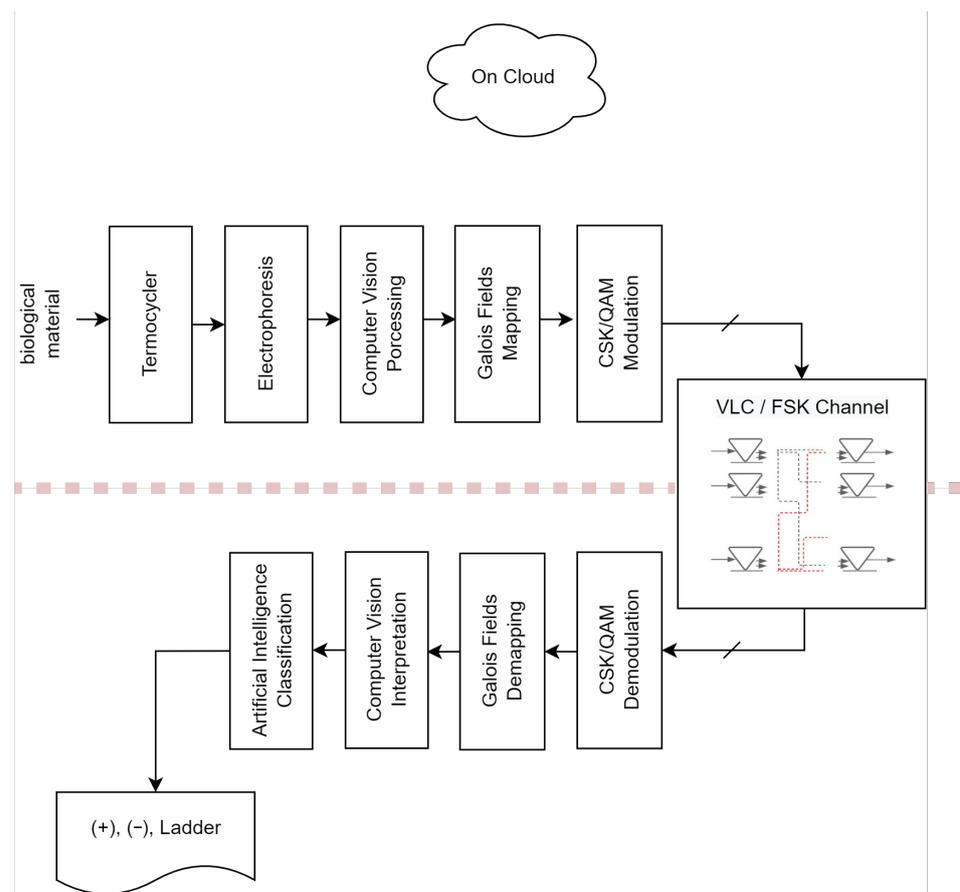


Figure 1. System diagram for the phases of model searching and operation of the classification model.

At the entrance of the second block, the resulting DNA samples are loaded into the second block, called “Electrophoresis”, which allows the generation of images of gels. The inverse block called “Artificial Intelligence Classification” represents the best model for classification.

In the third block, called “Computer Vision Processing”, the image from the previous block is filtered and delivered to the next block for numerical representation. On the receiving side, the reverse process is called “Computer vision interpretation”, where a numerical input is converted into an image for interpretation.

The remaining issues are associated with the communication process. Galois Field Mapping/Galois Fields Demapping is a function that converts images to polynomials in the emitter and polynomials to images in the receiver. In the fifth block, called “CSK/QAM modulation”, each of the numbers is calculated as the centroid of the CSK modulation and mapped into a QAM constellation, and these signals are sent over the MIMO channel. In the receiver, the block called “CSK/QAM demodulation” takes the QAM signals and converts them into numbers which are delivered to the Galois Fields Demapping block.

Figure 1 depicts a massive array of LEDs and photodetectors inside a box that represents the VLC/FSK MIMO channel. The segments connecting the antennas represent electromagnetic waves propagating in air molecules as photon jets or sine waves with an amplitude, frequency, and phase.

3.1. Line-of-Sight (LoS) Link

To obtain $h_{LoS}(t)$, the most fundamental VLC link is considered with a single light source (LS), which can be monochromatic or multichromatic, and a single PD in an indoor free-space environment. When considering the LS, a point source from the perspective of the PD, the optical received power $P_{R,opt}$ can be expressed as [71]:

$$P_{R,opt} = P_{T,opt} G_{conc} G_{filter} f(\theta, \theta_{1/2}) \frac{A_{R,eff}}{\pi r^2} \text{ for } d \gg \lambda \text{ and } A_R \gg \lambda^2, \quad (1)$$

where $P_{T,opt}$ is the optical transmission power, $G_{conc} \geq 1$ is the optical concentration gain, $G_{filter} \leq 1$ is the optical filtering loss, r is the distance between the LS and the PD, $\theta_{1/2}$ is the half-power angle of the light beam, A_R is the aperture area of the PD, and $A_{R,eff}$ is the effective aperture area of the PD such that [71]:

$$A_{R,eff} = A_R \cos \phi, \quad (2)$$

Note that the condition $d \gg \lambda$ stems from the point source assumption, while the condition $A_R \gg \lambda^2$ implies that the optical power detection process at the PD is deterministic. Note that when ϕ exceeds the field of vision (FOV) of the PD, ϕ_{POV} . Then, $P_{R,opt} = 0$ [71].

3.2. Non-Line-of-Sight (NLoS) Link

Multipath channels cause stochastic and time-varying signal distortion in Radio Frequency (RF) communications, causing microwave channels to be modelled as random. The multipath channel in the VLC, on the other hand, is deterministic because $A_R \gg \lambda^2$. In other words, the PD captures the optical signal over an area that is millions of times larger than a square wavelength. The indoor VLC channel is time-invariant as long as the objects in the room are fixed. Nonetheless, multipath propagation can cause intersymbol interference in VLC systems at high data rates, according to Hoeher [71]. To obtain $h_{NLoS}(t)$, it is easiest to begin with a single reflector. This reflector serves as a virtual light source (VLS). Because most reflections are diffuse, the angle of irradiance θ_2 is not always the same as the angle of incidence ϕ_1 . Furthermore, Lambertian reflections are commonly used. By using r_1 to represent the distance between LS and VLS and r_2 to represent the distance between VLS and PD, $G_{conc} G_{filter} = 1$, and Equation (3) can be extended to [71]:

$$P_{R,opt} = P_{T,opt} f(\theta_1, \theta_{1/2}) \frac{A_{ref,eff}}{\pi r_1^2} \cdot \rho(\lambda) f(\theta_2, 60^\circ) \frac{A_{R,eff}}{\pi r_2^2} \quad (3)$$

3.3. Transmitter: LED

The semiconductor light sources known as light-emitting diodes (LED) emit light when current flows through them. This is conceivable because of the electroluminescence phenomenon, in which the forward current causes semiconductor electrons to rejoin electron holes and release energy as photons. The energy required for electrons to traverse the band gap of the semiconductor determines the wavelength of the emitted light. Phosphor LEDs and Red Green Blue (RGB) LEDs are the two most commonly used types of LED to generate white light: (i) by using a blue LED with yellow phosphor, white light is produced and (ii) by using the RGB-based LED, which does not use phosphorus and is sustainable, high speeds can be reached in domestic environments [72].

3.4. Receiver

Photodiodes are photoelectric transducers because they convert signals of optical power to electrical impulses. The photodiode will generate an output current that is proportional to the incident optical power $R(\lambda)$ at a specific wavelength $\Phi(\lambda)$. It is measured in amperes per watt (A/W), as shown by Equation (4). The materials and structure of a photodiode determine its response curve.

$$R(\lambda) = \frac{i_r}{\Phi(\lambda)}, \quad (4)$$

3.5. The DC Gain of the Channel Model

When an optical input and optical output are considered, the Direct Current (DC) gain of the channel model is given by

$$H_{\text{ref, opt}} = H_{\text{LoS}} + H_{\text{NLoS,(1)}} + H_{\text{scatter}}, \quad (5)$$

where H_{LoS} is the LoS component, $H_{\text{NLoS,(1)}}$ is the single-hop NLoS component, and H_{scatter} is the NLoS contribution of light scattering off dust particles. Due to the insignificance of the following bounces in terms of the received power and time dispersion, only one hop is evaluated [25,37]. In Appendix A, we provide descriptions of reference model equations, which describe the most applicable reference models utilised for the UM-VLC Single input single output channel (UM-VLC SISO) [24,25]. In [38], it is assumed that particles are spread through a two-dimensional disc with irregular walls, nondeterministic diffuse reflections, shadows, and a scattering component.

4. Methodology

From right to left, Figure 1 shows the methodology, which comprises the MIMO channel, CSK/QAM Modulation and the DNA strand picture as a polynomial and vice-versa.

4.1. Model of a Channel Based on CSK/QAM Mapping

Due to its low cost and simplicity, intensity modulation with direct detection (IM/DD) is used in the majority of VLC systems. The transmitters in this kind of system are LEDs, and the instantaneous optical power $\Phi_e(t)$ is modulated in proportion to the driving electrical current $i_t(t)$, which is modulated in accordance with the data to be broadcast.

The optical power signal travels down the channel and eventually reaches the receiver's surface, which is often a photodiode, also known as a photodetector (PD). In the photodiode, the received optical power causes a proportionate photocurrent, $i_r(t)$. In this study, the Bit Error Rate (BER) was calculated using the Minimum Mean Square Error (MMSE) estimator for a variety of MIMO arrays [73]. In order to send the DNA samples that are in a sewer, an uplink process must be created. This assumes that there are N_t light sources capable of transmitting the signal and a receiving device with N_r Photo detector (PDs):

$$\mathbf{y} = \begin{pmatrix} h_{1,1} & \dots & h_{1,N_t} \\ \dots & \dots & \dots \\ h_{n_r,1} & \dots & h_{n_r,N_t} \end{pmatrix} \mathbf{x} + \mathbf{n}, \quad (6)$$

where \mathbf{y} is the $N_r \times 1$ vector representing the signal received at each PD, \mathbf{x} is the transmitted signal vector of size $N_t \times 1$, $h_{i,j}$ ($i = 1, \dots, N_r$ and $j = 1, \dots, N_t$) are the channel gain of the link between the j -th transmitter and the i -th PD, and \mathbf{n} is the $N_r \times 1$ vector representing the noise at each PD, including all possible noise, which can be expressed as Equation (5).

Furthermore, the mapping of an M-CSK constellation to an M-QAM constellation is proposed. Figure 2 represents an xyY diagram of the Commission Internationale de l'Éclairage (CIE) colour space from 1931. This was the first colour space based on experimental results of human colour perception. A colour space is a multidimensional collection of all colours that a certain colour model can generate. Historically, the IEEE 802.15.7 standard defines some guidelines for designing M-CSK constellations and directly applying them for modulation [74].

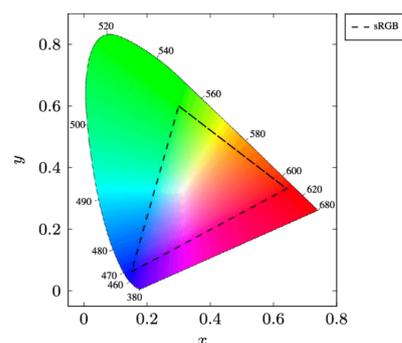


Figure 2. The xyY diagram of the Commission Internationale de l'Éclairage.

Figure 3 illustrates the fundamental principle of mapping for the chromosomal length N^{DNA} . Depending on the length of the DNA, $N^{DNA} = 2^{2^i}$ points in the M-CSK constellation are selected and mapped to the M-QAM constellation, which is depicted on the right side of Figure 3 as clouds of points, with a total of 2^{2^i} points $i = 0, 1, 2, 3 \dots$ shown.

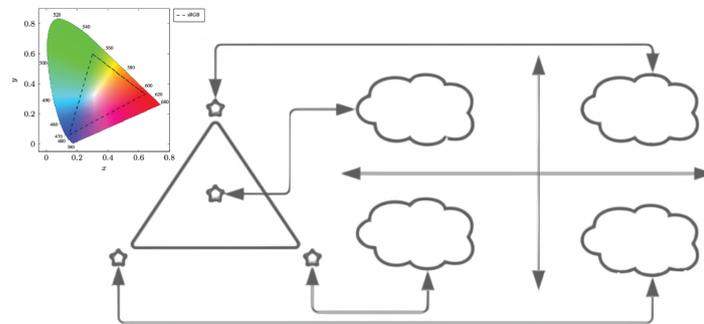


Figure 3. General CSK/QAM assignment.

Figure 4 represents 4-CSK constellation mapping into 4-QAM. Although points in the CIE 1931 xyY space can be assigned in any arbitrary order, the four 4-CSK stars correspond to the four 4-QAM stars. Due to their proximity, it is more difficult to pass separator hyperplanes to detect the four points of the 4-CSK constellation. On the right is a 4-QAM constellation with more evenly spaced stars, which makes it simpler to deliver separator planes to them.

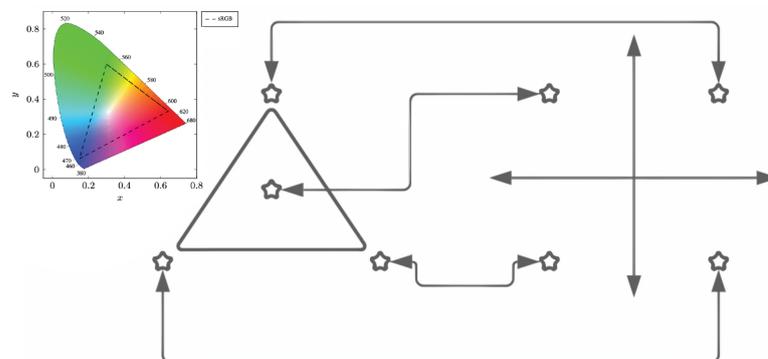


Figure 4. 4-CSK constellation mapping into 4-QAM.

The centre of the colour bands can be expressed in CIE 1931 xyY space coordinates as $\mathbf{s}_R, \mathbf{s}_G, \mathbf{s}_B \in \mathbb{R}^2$, which is known as the centre of band symbols. All colours that can be reproduced by LEDs via additive mixing form a triangle in CIE xyY space with the vertices $\mathbf{s}_R, \mathbf{s}_G$ and \mathbf{s}_B .

The gamut of the system is the set of all colours that may be reproduced by the three LEDs and is mathematically defined as the convex combination of the centre of band symbols in the CIE xyY space, as shown in

$$\mathcal{G} = \{a_1 \mathbf{s}_R + a_2 \mathbf{s}_G + a_3 \mathbf{s}_B \in \mathbb{R}^2\}; \quad a_1 + a_2 + a_3 = 1 \wedge \forall a_1, a_2, a_3 \geq 0 \quad (7)$$

where \mathcal{G} denotes the system's gamut.

The explicit CIE xyY coordinates of the symbol \mathbf{s}_k are denoted as $\mathbf{s}_k = [x_k, y_k]^T$. The symbol \mathbf{s}_k can be expressed as a radiant flux vector $\Phi = [\Phi_R, \Phi_G, \Phi_B]^T \in \mathbb{R}_{\geq 0}^3$, where Φ_R, Φ_G, Φ_B represent the radiant fluxes to be emitted by the red, green and blue LEDs respectively. The radiant flux vector Φ can be obtained by solving the system of Equation (8).

$$\begin{aligned}
x_k &= \Phi_R x_R + \Phi_G x_G + \Phi_B x_B \\
y_k &= \Phi_R y_R + \Phi_G y_G + \Phi_B y_B \\
1 &= \Phi_R + \Phi_G + \Phi_B
\end{aligned} \tag{8}$$

The centroid is determined by multiplying the four points $(x_k, y_k)(x_R, y_R)(x_G, y_G)(x_B, y_B)$ by their components and Φ , which enables the transmission of white light between the points. For separator planes, such as using a support vector machine, it is simpler to map these four points in a square constellation.

Using the cursors from the CSK, these points are mapped once more into a square regular M-QAM constellation. Data are processed by the M-QAM modulator, which then maps them into a plane with an in-phase and quadrature component.

$$s = a + jb \text{ where } a, b \in \{\pm 1, \pm 3, \dots, \pm(\lceil \sqrt{M} \rceil - 1)\} \tag{9}$$

Suppose we have k cursors according to Equation (A2), which indicates that all 2^{2^i} , $(i = 3)$ M-QAM elements of constellation have been addressed. The fastest method would be to undergo transmission in a single cycle, but this is unnecessary because transmission as a matrix $2^{2^i} \times 2^{2^i}$, $(i = 1)$ could occur, in which case the tables would be smaller according to \mathcal{G} .

Given a set of LEDs with 2^{2^i} colours with the same characteristics, except for having different Semiconductor Photo Detectors (SPDs), given by $\Phi_1(\lambda)$, $\Phi_2(\lambda) \dots \Phi_{2^{2^i}}(\lambda)$ respectively, spatially grouped so that their positions in space can be approximated from a sufficiently large distance and DPs, each with a spectral response $R(\lambda)$ and surface area A_R and filtered by an optical filter of one of the 2^{2^i} colours with spectral gains of $\mathcal{G}_1(\lambda)$, $\mathcal{G}_2(\lambda) \dots \mathcal{G}_{2^{2^i}}(\lambda)$, respectively, and spatially grouped in such a way that spatial positions can be approximated from a sufficient distance, if the distance d between the LEDs and PDs is large enough that a single emitter and receiver position is a reasonable approximation, then the gain of the UM-VLC DC electro-optical channel from the i -th LED to the j -th PD with $i, j = 1, 2, 3, \dots, 2^{2^i}$ can be given by:

$$H_{el}[i, j] = H_{LoS}^{el}[i, j] + H_{NLoS}^{el}[i, j], \tag{10}$$

where $H_{LoS}^{el}[i, j]$ is the DC gain of the LoS link, which can be expressed as:

$$H_{LoS}^{el}[i, j] = \frac{1}{P_{T,i}} \int R(\lambda) \mathcal{A}_{0,j}(\lambda) \Phi_i(\lambda) d\lambda \tag{11}$$

$$= \mathcal{C} \frac{(m+1)A_R}{2\pi d^2 P_{T,i}} \Omega_{LoS} \Psi_{LoS} \int R(\lambda) \mathcal{G}_j(\lambda) \zeta_0(\lambda) \Phi_i(\lambda) d\lambda, \tag{12}$$

where $P_{T,i} = \int \Phi_i(\lambda) d\lambda$. Similarly, $H_{NLoS}^{el}[i, j]$ is the DC gain of the single-hop NLoS link, which can be expressed as:

$$H_{NLoS}^{el}[i, j] = \sum_{w=1}^W H_{NLoS,w}^{el}[i, j] \tag{13}$$

$$= \frac{1}{P_{T,i}} \sum_{w=1}^W \int R(\lambda) \mathcal{A}_{w,j}(\lambda) \Phi_i(\lambda) d\lambda \tag{14}$$

$$= \frac{(m+1)A_R}{2\pi P_{T,i}} \sum_{w=1}^W \mathcal{C}_w \frac{A_{refl,w}}{d_{1,w}^2 d_{2,w}^2} \Omega_{NLoS,i}^{(w)} \Psi_{NLoS}^{(w)} \tag{15}$$

$$\int R(\lambda) \rho_w(\lambda) \mathcal{G}_{w,j}(\lambda) \zeta_w(\lambda) \Phi_i(\lambda) d\lambda \tag{16}$$

$$\begin{aligned}
 &= \frac{(m+1)A_R}{2\pi P_{T,i}} \sum_{w=1}^W C_w \frac{A_{\text{refl},w}}{d_{1,w}^2 d_{2,w}^2} \Omega_{\text{NLoS},i}^{(w)} \Psi_{\text{NLoS}}^{(w)} \times \\
 &= \sum_{\lambda=1}^{\lambda=2^{2^i}} C_w R(\lambda) \rho_w(\lambda) \mathcal{G}_{w,j}(\lambda) \zeta_w(\lambda) \Phi_i(\lambda),
 \end{aligned} \tag{17}$$

It should be noted that the integral is discretised and that 2^{2^i} points are taken into account. If it is considered that the optical filters do not depend on the angle of incidence θ_w , the definition can be simplified to $\mathcal{G}_{w,j}(\lambda) \equiv \mathcal{G}_j(\lambda)$. $\Omega_{\text{NLoS},i}^{(w)}$ depends on the i -th LED, because light bouncing on the w -th reflector will have different angles. This is to account for the irregularity of the underground tunnel walls, where a small shift in the LS can have a large effect on the angle of reflection off the wall. Because the LEDs are slightly separated from one another, light from different LEDs will bounce off the walls independently. The channel impulse response for the UM-VLC channel between the i -th LED and the j -th PD is then given by [71]:

$$h[i,j](t) = H_{\text{LoS}}^{\text{el}}[i,j] \cdot \delta\left(t - \frac{d}{c}\right) + \sum_{w=1}^W H_{\text{NLoS},w}^{\text{cl}}[i,j] \cdot \delta\left(t - \frac{d_{1,w} + d_{2,w}}{c}\right), \tag{18}$$

Given an optical power signal $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_{2^{2^i}}(t)]^T$ (W) as input, where $x_i(t)$ is the optical power signal emitted by the i -th LED, for $i = 1, 2, \dots, 2^{2^i}$, then the received photocurrent signal at the j -th PD, $y_j(t)$ (A) can be given by:

$$y_j(t) = \sum_{i=1}^{2^{2^i}} h[i,j](t) * x_i(t) + n_j(t), \tag{19}$$

where $n_j(t)$ is the noise at the j -th PD with noise variance of σ_j^2 .

4.2. Galois Field Mapping/Galois Fields Demapping

This section describe how to convert a dsDNA image into a polynomial and vice versa as well as how to colour-code strands or dsDNA fragments using M-CSK/M-QAM modulation. A chromosome contains a single long molecule of DNA, only part of which corresponds to an individual gene. We developed a simple DNA-based model to represent the fields $GF(p)$ and $GF(p^n)$, $n > 1$. It is based on the differential migration of dsDNA fragments of different sizes in gel electrophoresis, which is a standard technique for dsDNA fragments of different sizes that have previously been obtained by PCR. Here, the size of a dsDNA fragment corresponds to the number of base pairs [bp] that are contained in the fragment.

Each element $r \in GF(p)$ is represented by a dsDNA fragment whose size is unique to the element r . Therefore, only p dsDNA fragments are necessary to represent all elements of $GF(p)$. Table 1 shows this representation using dsDNA fragments of different sizes, where the smallest size S_0 is composed of one or more genes and the largest is S_{p-1} .

Table 1. DNA representation for elements in $GF(p)$.

$r \in GF(p)$	0	1	2	...	$p - 1$
Size of DNA Fragment [bp]	S_0	S_1	S_2		S_{p-1}

Gel electrophoresis is used to visualize the DNA molecular representation of a nonzero element $\alpha^k \in GF(p^n)$, which represents the coefficients of the polynomial expression given for Equation (20), as shown in Table 2.

Table 2. dsDNA fragment representation of α^k performed by agarose gel electrophoresis.

Referential DNA Site	Positions of the Coefficients $\alpha^k \in GF(p)$					
	$n - 1$	$n - 2$...	2	1	0
S_{p-1}		x			x	x
...						
S_3			...			
S_2	x					
S_1				x		
S_0						

The dsDNA fragments for each coefficient $a_i \in GF(p), i = 0, 1, \dots, n - 1$ are loaded into different slots of the agarose gel matrix. The slots and their respective columns are numbered $n - 1, n - 2, \dots, 2, 1, 0$ according to the order of powers $\alpha^{n-1}, \dots, \alpha^2, \alpha, 1$ from left to right. Then, an electric field is applied to force the molecules to migrate through the gel and be separated by size.

$$\alpha^k = 2\alpha^{n-1} + (p-1)\alpha^{n-2} + \dots + \alpha^2 + (p-1)\alpha + (p-1). \quad (20)$$

For this purpose, chains of size S_2 were loaded into slot $n - 1$, chains of size S_{p-1} were loaded into slot $n - 2$, and from slot $n - 3$ to slot 3, chains of size S_0 were loaded. Finally, chains of size S_1 were loaded into slot 2, and chains of size S_{p-1} were loaded into slots 1 and 0. Thus, our model defines a unique DNA-based representation for each element of $GF(p^n)$.

We should note that $\alpha^0 = 1$, and the null element $0 \in GF(p^n)$ does not have a representation as a power of α . Hence, the field $GF(p^n)$ has p^n elements, which are stored in a lookup table according to the power of each element.

Example 1. To construct the field $GF(2^3)$, $n = 3$ a new element α is added to the field $GF(2)$. α is a root of the primitive polynomial $P(x) = x^3 + x + 1$ with a degree of $n = 3$, which is used to generate the elements of $GF(2^3)$. Since α is a root of the polynomial $P(\alpha) = \alpha^3 + \alpha + 1 = 0$ then, $\alpha^3 = \alpha + 1, \alpha^4 = \alpha\alpha^3 = \alpha(\alpha + 1)$ and so on. The field $GF(2^3)$ has $2^3 = 8$ elements.

In Table 3, a $\alpha = \alpha^1$ is introduced as a root, but $\alpha^0 = 1$ and 0 can also be introduced, since they do have representation in the field. In the case of a $GF(2^{16})$ field containing $2^{16} = 256$ elements, the same method can be utilised to organise the elements for use in 16x16 MIMO arrays.

Table 3. Generation of the polynomials from $P(x) = x^3 + x + 1$.

Element $GF(2^3)$	Polynomial	Symbol
0	0	000
α^0	1	001
α^1	α	010
α^2	α^2	100
α^3	$\alpha + 1$	011
α^4	$\alpha^2 + \alpha$	110
α^5	$\alpha^2 + \alpha + 1$	111
α^6	$\alpha^2 + 1$	101

4.3. AI-Based Procedures to Replace Human Processing

In Section 4.2, the interaction between Artificial Intelligence algorithms and modulation/demodulation was described. At both the transmitter and receiver, the Scikit-learn Python module is employed [75].

4.3.1. Logistic Regression

The weighted sum of the input attributes is used in logistic and linear regressions. However, the logistic regression bias has a binary output as opposed to a direct output. According to Suykens [76], a logistic regression model predicts that if the probability is less than 50%, it belongs to the negative class denoted by “A” or “0”, and if it is greater, it belongs to the positive class denoted by “B” or “1”.

To find the value of the prediction, Equation (21) can be used:

$$\frac{\partial}{\partial \theta_j} MSE(\theta) = \frac{2}{m} \sum_{i=1}^m (\theta^T \cdot \mathbf{x}^{(i)} - y^{(i)}) x_j^{(i)}, \quad (21)$$

where m is the number of partial derivatives, \mathbf{x} is the input, and y is the predicted value. Equation (22) represents the logistic regression model’s estimated probability in vector form \hat{p} :

$$\hat{p} = h_{\theta}(x) = \sigma(\theta^T \cdot x), \quad (22)$$

where θ is the vector of the model parameters, θ^T is the transpose of θ , h_{θ} is the hypothesis function, and $\sigma(\cdot)$, a logistic or logit sigmoidal function, generates a number between 0 and 1, as shown in Equation (23).

$$\sigma = \frac{1}{1 + \exp(-t)} \quad (23)$$

After estimating the probability $\hat{p} = h_{\theta}(x)$ that an instance x belongs to the positive class, the Logistic Regression model can easily make its prediction \hat{y} . The logistic regression model’s prediction is shown in Equation (24).

$$\hat{y} = \begin{cases} 0 & \text{if } \hat{p} < 0.5 \\ 1 & \text{if } \hat{p} \geq 0.5 \end{cases} \quad (24)$$

Note that $\sigma(t) < 0.5$ when $t < 0$, and $\sigma \geq 0.5$ when $t \geq 0$, so a logistic regression model predicts 1 if $\theta^T \cdot x$ is positive and 0 if is negative.

4.3.2. Naive Bayesian with Gaussian optimisation

The Naive Bayesian with Gaussian optimisation (GaussianNB) method finds promising parameter values by using a Gaussian process model of the objective function [77]. The Probability of Improvement (PI) is an intuitive strategy that can be calculated analytically by using Gaussian processes to maximise the probability of improvement over the best current value [78]:

$$a_{PI}(x; \{x_n, y_n\}, \theta) = \Phi(\gamma(x)), \quad (25)$$

$$\gamma(x) = \frac{f(x_{best}) - \mu(x; \{x_n, y_n\}, \theta)}{\sigma(x; \{x_n, y_n\}, \theta)}, \quad (26)$$

where $f : \chi \rightarrow \mathbb{R}$, x_{best} is the best current value, μ is its predictive mean function and σ is the predictive variance function. In order to maximise the expected improvement over the best current value, the Expected Improvement (EI) could also be calculated using a Gaussian process:

$$a_{EI}(x; \{x_n, y_n\}, \theta) = \sigma(x; \{x_n, y_n\}, \theta) (\gamma(x) \Phi(\gamma(x)) + \mathcal{N}(\gamma(x); 0, 1)), \quad (27)$$

where a_{EI} represents the acquisition function with the highest expected improvement, $\Phi(\cdot)$ represents the cumulative distribution function and \mathcal{N} represents the normal distribution. The upper confidence limit of Gaussian Processes (GP) seeks to exploit the concept of lower

and upper confidence limits in the maximisation case in order to build acquisition functions that minimise regret as optimisation progresses [79]:

$$a_{LCB}(x; \{x_n, y_n\}, \theta) = \mu(x; \{x_n, y_n\}, \theta) - \kappa\sigma(x; \{x_n, y_n\}, \theta), \quad (28)$$

where $a_{LCB}: \mathcal{X} \rightarrow \mathbb{R}^+$ denotes the acquisition function, LCB is the lower confidence bound and κ is tunable to balance exploitation versus exploration.

4.3.3. SVM Classifier

A support vector machine divides the elements of a set into different subsets known as classes with the goal of finding the widest possible hyperplane that best separates these classes. The margin can be seen in Figure 5. It is defined as the maximum width of the region parallel to the hyperplane that has no interior data points. Equation (29) shows how a linear SVM predicts the class of a new x instance by calculating the decision function $w^T x + b = w_1 x_1 + \dots + w_n x_n + b$: if the result is positive, the predicted class $f(x)$ is the positive class (1); otherwise, it is the negative class (0) [75]. b is the bias and w is the feature weight.

$$f(x) = \begin{cases} 0 & \text{si } w^T x + b < 0 \\ 1 & \text{si } w^T x + b \geq 0 \end{cases} \quad (29)$$

To make it easier to separate the classes after this transformation, kernel functions move the data to a different, usually higher, dimensional space, potentially simplifying nonlinear complex decision boundaries in the assigned higher dimensional feature space to make them linear. The data do not have to be explicitly transformed in this process, which is known as a kernel trick [80]. A second-degree polynomial kernel is the function $K(a, b) = (a^T b)^2$. Based on some mapping ϕ , the kernel K corresponds to an inner product in a feature space [81]. A kernel is a function in ML that computes the dot product $\phi(a)^T \phi(b)$ by using only the original vectors \mathbf{a} and \mathbf{b} without computing the ϕ transformation. The polynomial kernel for polynomials of degree d is shown in Equation (30) [75].

$$K(a, b) = (\gamma a^T b + r)^d \quad (30)$$

where a and b are vectors in the input space, $r \geq 0$ is a free parameter that compensates for the impact of higher-order terms in the polynomial versus lower-order terms in the polynomial, and γ is a scaling parameter.

When $r = 0$, it is said that the kernel is homogeneous. When $d = 1$ and $r = 0$ are implemented, the result is identical to that of a linear kernel. If d is greater than one, nonlinear decision limits are produced, with the degree of nonlinearity increasing as d increases. Due to overfitting, d values greater than 5 are typically not recommended. Figure 5 depicts the optimal hyperplane with a polynomial kernel separating the data, where the light blue and brown dots represent data belonging to two distinct classes. The segmented red lines represent the various hyperplanes that can be constructed to partition data representing two classes between two point clouds. In a similar fashion, the red line represents the hyperplane that maximises class separability.

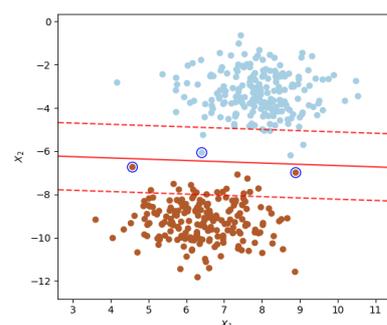


Figure 5. Graph of a polynomial SVM showing the hyperplane separating the samples from the classes.

4.3.4. Extra Trees Classifier

The Extra Trees Classifier (ETC), also known as extremely random trees, generates a large number of decision trees, but the per-tree sampling is random. Tanha [82] used this method to assemble a data set with unique samples in each tree. According to Geurts [83], the geometric analysis generated by the ETC algorithm assumes a minimum number of samples ($n_{min} = 2$). When the number of trees is $M \rightarrow \infty$, the models generated by the Extra Trees algorithm appear to be linear. Thus, with the minimum sample condition $n_{min} \geq 2$, the algorithm can be extrapolated for the n-dimensional case. In this way, a continuous multilinear approximation is obtained for the case of infinite samples $N \rightarrow \infty$. In either case, the expression presented in Equation (31) can be used, where $x_i = (x_{1,i}, \dots, x_{n,i})$ is a n dimensional input vector, yielding y_i as the output. To simplify the notation, we give the notation presented in the Equation (32), where j th indicates the value of the sample, so that $\forall (i_1, \dots, i_n) \in \{0, \dots, N\}^n$. For $I_{(i_1, \dots, i_n)}(x)$, the characteristic function of the hyperplane corresponds to the one presented in Equation (33).

$$ls_N = \{(x_i, y_i) : i = 1, \dots, N\} \tag{31}$$

$$x_{j,(0)} = -\infty \text{ and } x_{j,(N+1)} = +\infty, \forall j = 1, \dots, n \tag{32}$$

$$[x_{1,i_1}, x_{1,(i_1+1)}] \times \dots \times [x_{n,i_n}, x_{n,(i_n+1)}] \tag{33}$$

This enables us to demonstrate that, as stated in Zhao [84], an infinite number of extra trees will generate an approximation of the form presented in Equation (34). Thus, for fully developed trees, the development shown in Equation (35) is shown.

$$\hat{y}(x) = \sum_{i_1=0}^N \dots \sum_{i_n=0}^N I_{(i_1, \dots, i_n)}(x) \sum_{X \subset \{x_1, \dots, x_n\}} \lambda_{(i_1, \dots, i_n)} \prod_{x_j \in X} x_j \tag{34}$$

$$\hat{y}(x_i) = y_i, \forall (x_i, y_i) \in ls \tag{35}$$

A piecewise linear model is obtained for the specific case of a one-dimensional input, as shown in Equation (36), where $I_{(i)}(x_1)$ is the interval characteristic function, and the values of $\lambda_{i,\phi}$ and λ_{i,x_1} are obtained from Equations (35) and (36).

$$\hat{y}(x) = \sum_{i_1=0}^N I_{(i_1)}(x) \sum_{X \subset \{x_1\}} \lambda_{(i_1),X} \prod_{x_j \in X} x_j = \sum_{i=0}^N I_{(i)}(x_1) (\lambda_{i,\phi} + \lambda_{j,s\{x_1\}} x_1) \tag{36}$$

4.3.5. Histogram Gradient Boosting Classifier

Decision trees also inspired gradient boosting, one of the most useful algorithms for generating table structures and enabling predictive regression modelling, according to Padhi [85]. There are two variants that are based on the operating system implementation: Light Gradient Boosting (LGB) and GPU-accelerated XGBoost. LGB is a fast, distributed, high-performance gradient boosting framework based on the decision tree algorithm that can be used for ranking, classification and a variety of other ML tasks [86,87]. This model reduces the learning process time by at least 20 times while maintaining the same precision. According to Chen [88], the XGBoost algorithm boosts the GPU performance by using perfect shuffling of indexes and data in parallel sums and GPU-accelerated sorting, generating trees of all data concurrently for each iteration.

XGBoost is an enhanced version of the gradient boosting algorithm that is more efficient and scalable. Automatic feature extraction is one of the characteristics that distinguishes XGBoost from other algorithms. XGBoost supports regularisation to prevent overfitting and has the capacity to learn from nonlinear datasets. In addition, the parallelisation feature enables XGBoost to utilise multiple CPU cores. It is one of the tree-based

additive ensemble models that consists of a group of base learners. XGBoost can generally be represented by:

$$F = (m_1, m_2, m_3, m_4 \dots m_n), \hat{y}_i = \sum_{t=1}^n m_t(x_i) \tag{37}$$

where \hat{y}_i is the final predictive model, which is the combination of all weak learners, and x is the input feature for each weak learner, i.e., m .

From the paper [87], we extracted the objective function for XGBoost, as given below:

$$\text{Obj}(\theta) = \sum_{i=1}^m L(z_i, z_i) + \sum_{t=1}^T \Omega(f_t). \tag{38}$$

In Equation (38), note that the objective function has two parts; the first part denotes the loss function, i.e., L denotes the training loss of either the logistic or squared loss, and the second part represents the addition of each tree's complexity. z_i is the actual value and \hat{z}_i is the predicted value, whereas Ω is the regularisation term, T denotes the total number of trees, and f is the function.

4.3.6. Model Evaluation

Figure 6 depicts a classification table displaying the various error types.

		REAL	
		POSITIVE	NEGATIVE
PREDICTED	POSITIVE	TP	FP
	NEGATIVE	FN	TN

Figure 6. Outputs of the classification model.

The formula for calculating the *recall* parameter, which is relevant for the assessment of type 2 errors or false negatives, is presented in Equation (39):

$$\text{recall} = \frac{TP}{TP + FN}, \tag{39}$$

where TP stands for True Positives and FN stands for False Negatives [89].

5. Results Analysis

The interface illustrated in Figure 7 corresponds to Galois Field Mapping/Galois Fields Demapping. During the phase of finding the ideal algorithm, the transmitter converts images to polynomials (shown from left to right), while the receiver converts polynomials to images (shown from right to left), as displayed in Figure 7. The time has come to reveal the findings. First, the results of the channel based on CSK/QAM mapping will be shown, followed by the SARS-CoV-2 Searching results of the model, and finally, the SARS-CoV-2 Operation results of the best model.

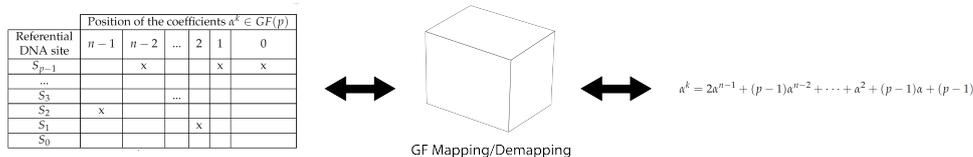


Figure 7. Galois Field Mapping/Galois Fields Demapping procedure.

5.1. Results of the Channel Based on CSK/QAM Mapping

The communications channel is the source of all negative effects when collecting data for the purpose of locating and implementing the optimal model. Traditionally, IEEE 802.15.7 specifies rules for designing M-CSK constellations and directly applying them for modulation. However, the mapping defined in Section 4.1 can also be used to indicate which cursor to map. Then, use a square constellation with better separation properties can be used. Figures 8 and 9 illustrate the experimental results for $N = 2^{2^i} \times 2^{2^i}$, $i = 0, 1$. Despite their two-dimensional representation in the CIE xyY plane, the null element points can be aggregated with an integer cursor can be assigned to each of them, and a table can be created in the cloud to provide copies at the transmitter and receiver for calculating inverse mapping between the M-CSK and the M-QAM constellation.

Figure 10 shows the MIMO channel capacity for $N = 2^{2^i} \times 2^{2^i}$ ($i = 0, 1, 2, 3$). In order to get a greater spectrum efficiency for a longer chromosome, it has been demonstrated that square constellations should be favoured due to their superior separation properties, despite their exponentially increasing complexity.

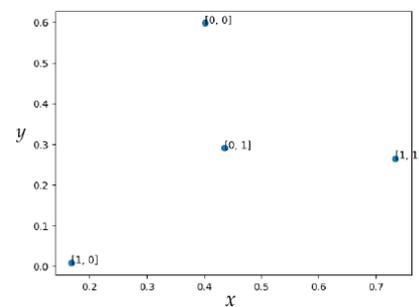


Figure 8. CSK constellation for $N = 2^{2^i} \times 2^{2^i}$, $i = 0$.

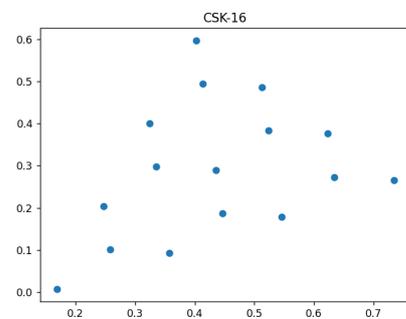


Figure 9. CSK constellation for $N = 2^{2^i} \times 2^{2^i}$, $i = 1$.

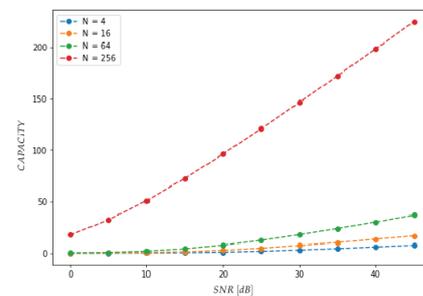


Figure 10. MIMO channel capacity for $N = 2^{2^i} \times 2^{2^i}$ ($i = 0, 1, 2, 3$).

It is feasible to transmit monochromatic photon streams. When coherent monochromatic frequencies are employed, energy is not wasted on phase incoherence effects, which ordinarily result in self-destructive phase effects. Consequently, the outcomes are improved.

Due to the fact that quantum objects do not require their attributes to be specified, a beam of baseband light can arrive at the photodetector from a variety of angles other than a coherent straight line. The use of a laser decreases costs because white light amplification equipment is avoided due to the high concentration of energy in a single frequency tone.

In order to generalise the FSK channel to a MIMO channel [51], Figures 11 and 12 show the output of the MIMO demodulator that will enter the “Galois Field demapping” process for 256-point QAM square constellation with SNR values of $SNR = 60[dB]$ and $SNR = 30[dB]$.

Figure 13 compares CSK/QAM mapping to QAM mapping by using the XGBoost algorithm in terms of the BER for different SNR values with $N = 2^{2^i} \times 2^{2^i}$, $i = 0, 1, 2, 3$. This simulation was carried out in 1[dB] steps for values up to $SNR = 80[dB]$. The BER was computed within one cycle using the MMSE estimator. This procedure was repeated 10,000 times to accumulate the erroneous values for each of the $SNR[dB]$ levels in an array. In contrast to linear mapping, the combination of M-CSK and M-QAM mapping results in nonlinear productions, i.e., it breaks the regularity of selecting the same points due to the centroid calculation and the inclusion of points as the null element. As mentioned in Section 4.3.5, the XGBoost algorithm produces the best results because it manages to generate multiple trees and is the only one capable of learning a nonlinear dataset, because it generates a new objective function. When the dataset is small, it cannot learn to predict the values to come, for example, when $N = 2^{2^i} \times 2^{2^i}$, and $i = 0, 1$ points occur, although it is seen to start working for $N = 2^{2^i}$, $i = 3$, but when the dataset increases, the gain is significantly improved. For $BER=10^{-3}$, gains of -10 [dB], -3 [dB], 3 [dB] and 5 [dB] occur for $N = 2^{2^i} \times 2^{2^i}$, $i = 0, 1, 2, 3$, respectively. It is concluded that the square constellation $N = 2^{2^i} \times 2^{2^i}$, $i = 3$ produces a greater benefit. When the data set is small, the algorithm is unable to learn to forecast future values. Alternatively, the BER improves as the data set is enlarged, making it simpler to separate states and creating a larger forest.

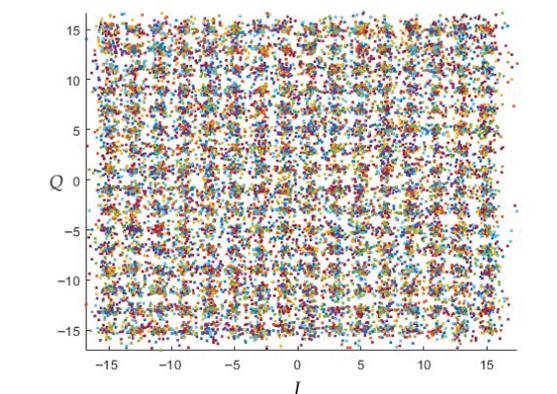


Figure 11. QAM constellation for $N = 2^{2^i} \times 2^{2^i}$, $i = 3$ on the MIMO channel, $SNR = 60[dB]$.

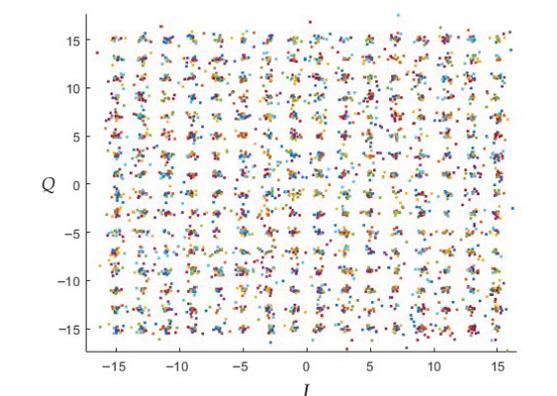


Figure 12. QAM constellation for $N = 2^{2^i} \times 2^{2^i}$, $i = 3$ on the MIMO channel, $SNR = 30[dB]$.

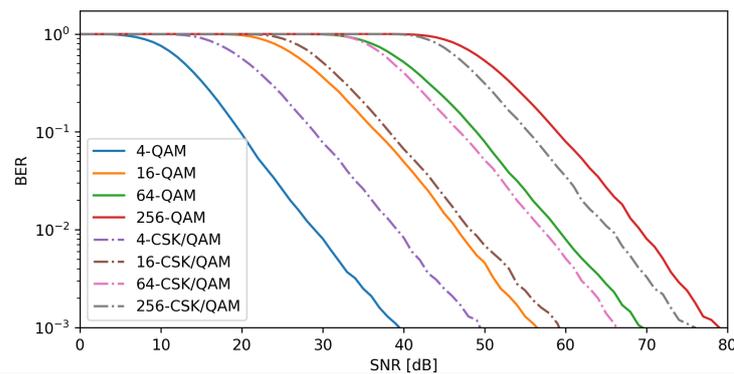


Figure 13. BER at different SNR levels in a $N = 2^{2^i} \times 2^{2^i}$ MIMO array for $i = 0, 1, 2, 3$.

5.2. Results SARS-CoV-2 Searching of the Model

Based on Section 4, this subsection analyses the proposed strategies in order to gather the information required for model searching and the operation of the best model from Section 5.3. The obtained biological material is subjected to the thermocycling process depicted in Figure 1, and the results are subsequently deposited on the electrophoresis gel, which is imaged after the reaction occurs. This is illustrated in Figure 14.

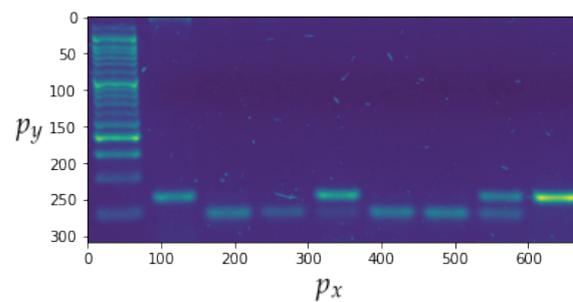


Figure 14. Image of the electrophoresis gel where p_x is the position of the horizontal pixel and p_y the position of the vertical pixel.

In addition to the channel noise described previously, the sample includes a significant amount of background noise. The use of a denoising convolutional auto-encoder model contributes to enhancement of the sample quality [32]. Figure 15 presents a comparison of the input and output images, with the output image containing less background noise. The detection of bands is the second step in the image processing procedure. Figure 16 depicts the outcomes of applying the methodology. The bands can be segmented with the data that will be subsequently analysed.

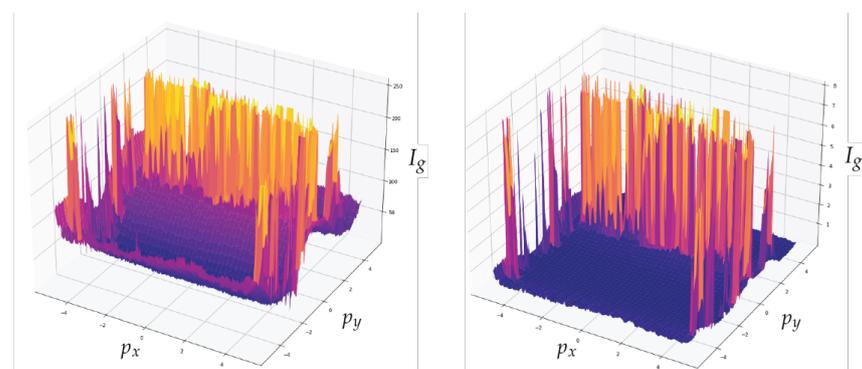


Figure 15. Comparison between the input image and the output of the convolutional denoising autoencoder model, where p_x is the position of the horizontal pixel, p_y is the position of the vertical pixel and I_g is the grayscale intensity of the pixel.

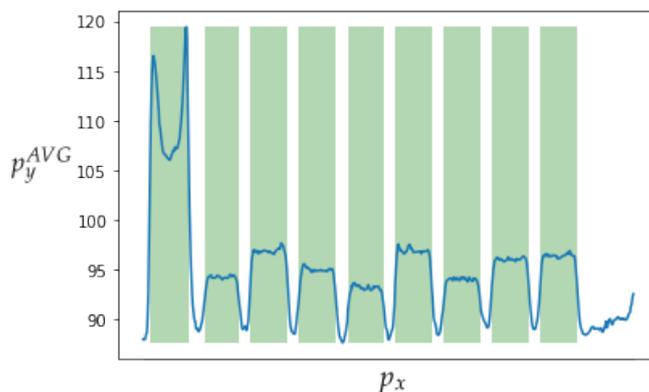


Figure 16. Detection of bands in an electrophoresis sample where p_y^{AVG} is the average value of p_y and p_x is the horizontal value of the pixel.

Figure 17 depicts data obtained from one of the bands. The top image shows the automatic clipping of the ladder band and the bottom image shows the average curve. Figure 18 show the recognition of peaks in the average curve obtained from the bottom image in Figure 17. This allows numerical representation through the method discussed in Section 4.2. Figures 16 and 17 show the average bands p_x^{AVG} and p_y^{AVG} . This information is used to train various mathematical models that enable band classification from electrophoresis bands. Figure 19 depicts the outcome of applying the Pearson correlation between the bands, demonstrating how the correlation of “band0” to the other bands is too low in comparison with the other values, which are higher than 0.5%. This produces an accuracy of 100%. Unfortunately, it cannot be used for the classification of positive and negative samples due to the high level of error.

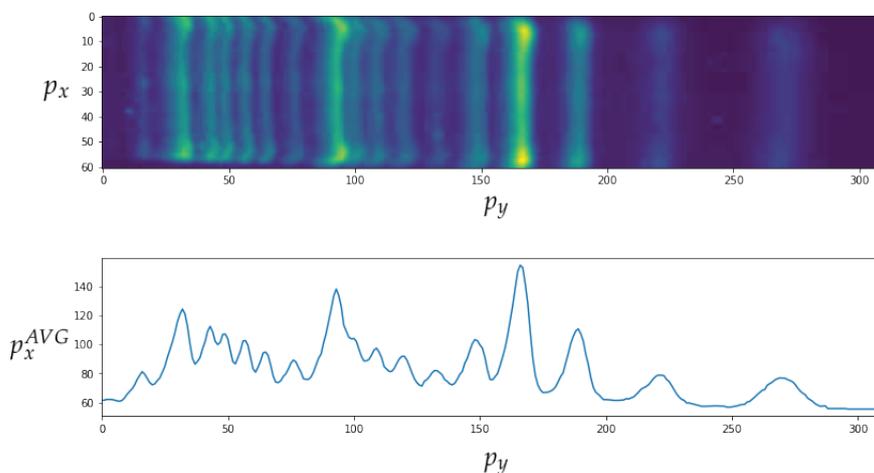


Figure 17. The top image shows the automatic clipping of the ladder band. The bottom image shows the average curve, where p_y is the vertical position and p_x^{AVG} is the average horizontal value p_x .

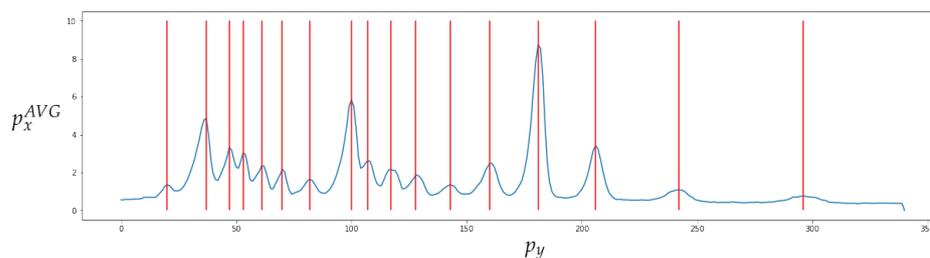


Figure 18. Recognition of peaks in the average curve obtained from the bottom image in Figure 17.

Based on these data, classification models can be trained to differentiate between three categories: "Positive (+)," "Negative (-)" and "Ladder". The training results of the models described in Section 4.3 are shown in Figure 20. In Table 4, it can be seen that the XGBoost classifier has the highest training accuracy of 96.03% compared with the other models and a *recall* rate of 99% for positive values.

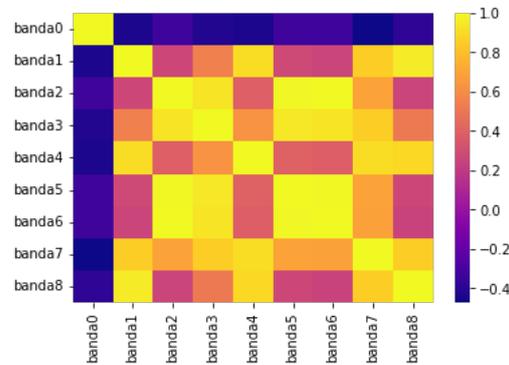


Figure 19. Application of Pearson’s correlation on the different columns of the electrophoresis sample.

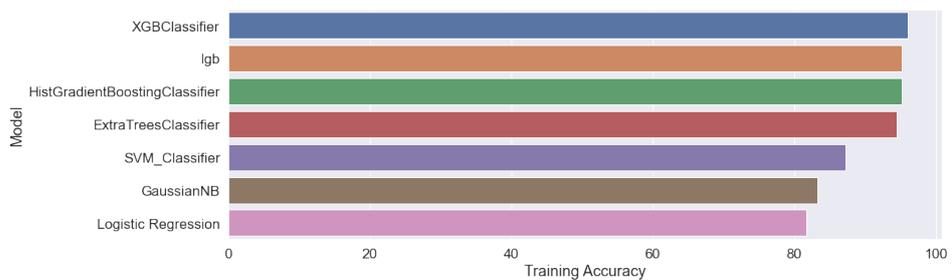


Figure 20. Comparison of the accuracy levels of the different trained classification models.

Table 4. Parameters obtained from the training of the XGBoost classifier.

Training Accuracy	100.0%			
Model Accuracy Score	96.03%			
Classification Report				
	precision	recall	f1-score	support
Ladder	1.00	1.00	1.00	17
Positive (+)	0.95	0.99	0.97	85
Negative (-)	0.95	0.83	0.89	24
Accuracy			0.96	126
Macro avg	0.97	0.94	0.95	126
Weighted avg	0.96	0.96	0.96	126

5.3. Results SARS-CoV-2 Operation of the Best Model

ML involves the parallel calculation of all processes and the selection of the best one; however, the processes may be conducted sequentially depending on the computer available. Images may be presented to Galois Fields mapping or switched directly to CSK/QAM modulation, depending on how the operation is configured, by entering the cursor *k* from Equation (A2).

In this instance, the most important parameter to investigate is *recall*, which indicates how frequently the model generates type 2 errors. In terms of both this metric and precision, the XGBoost classifier model has the best performance. The associated parameters are

displayed in Table 4. It can be seen that the model classifies the ladder correctly in all instances. It is important to note that *recall* is computed using Equation (39). Figure 14 depicts the results of the application of this method to the image shown in Figure 21. This corresponds to the first three boxes in the transmitter and the last three boxes in the receiver in Figure 1, which were previously completed manually but are now performed by an ML subsystem.

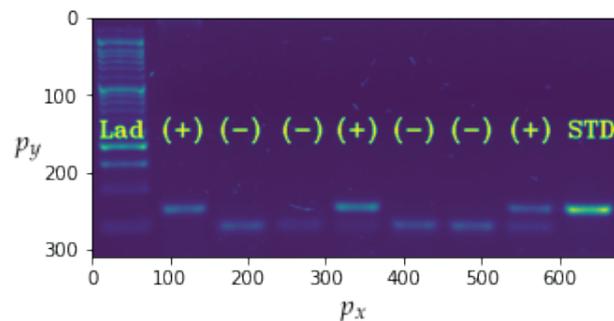


Figure 21. Results of the application of XGBoosts.

The accuracy of trained models will always be determined by the criteria used by the medical professional who prepared the data set. Due to uncertainty, the data can be propagated to the VLC channel. At the time of operation, the developed system is merely a tool; the results must be confirmed and interpreted by another health professional to determine the presence or absence of SARS-CoV-2. The created system reduces sample recognition times, allowing professionals to make more accurate diagnoses, and expands the data set size.

6. Conclusions

In this research work, an innovative VLC-based method for detecting COVID-19 in a subterranean environment was proposed. It was found that the unfavourable effects of the underground channel on VLC communications can be mitigated through precise mathematical modelling of the underground channel.

In order to get a higher spectrum efficiency for longer chromosomes, it has been shown that square constellations should be favoured due to their superior separation qualities within a photon stream. Transmission of monochromatic photon streams is an additional alternative. Since no energy is expended on phase incoherence effects, which generally result in self-destructive phase effects, when coherent monochromatic frequencies are employed, the results improve. It was also established that employing a laser saves money because there is no need for white light amplification equipment, often known as a driver. In addition, it was revealed that mathematical scaffolding in the exponential representation of DNA, in conjunction with the novel modulation and suitable channel modelling, prevents the transmission of heavy images.

The XGBoost technique was found to be the most successful, since it generates a large number of trees and is the only one that can learn a nonlinear data set by creating a novel goal function. When $N = 2^{2^i} \times 2^{2^i}$, $i = 0, 1$ points are used. For example, the dataset is too small for the algorithm to learn to predict future values, despite a slight improvement when $N = 2^{2^i} \times 2^{2^i}$, $i = 1$ points are used. As the dataset expands in size, the gain increases dramatically; this is something that linear models cannot achieve. For $\text{BER} = 10^{-3}$, gains of -10 [dB], -3 [dB], 3 [dB] and 5 [dB] were achieved for $N = 2^{2^i} \times 2^{2^i}$, $i = 0, 1, 2, 3$, respectively. The conclusion is that the square constellation $N = 2^{2^i} \times 2^{2^i}$, $i = 3$ yields a greater profit. During the searching phase, a classification algorithm was selected from a pool of available options. For a total of 630 COVID-19 samples, the best model was XGBoosts, which displayed an accuracy of 96.03% and a *recall* rate of 99% for positive values, placing its performance above that of the other models.

Furthermore, the uncertainty in the data propagates to the channel, so the accuracy of the trained models is determined by the criteria employed by the expert who creates the dataset. Clearly, the only way to rectify this is to compile a dataset from multiple sources so that it is complete and objective.

By extracting genetic information more efficiently, it is possible to classify the bands present in electrophoresis samples by using ML and a three-state classification process to determine whether the band corresponds to a COVID-19 positive, negative or ladder sample.

Author Contributions: Conceptualisation, I.S. and R.Z.-I.; methodology, I.S. and D.Z.-B.; software, R.Z.-I.; validation, I.S., W.A. and V.G.; formal analysis, I.S., R.B. and P.P.J.; investigation, I.S. and R.Z.-I.; resources, I.S.; data curation, V.G.; writing—original draft preparation, I.S., W.A., V.G., R.B. and M.I.; writing—review and editing, I.S., R.Z.-I. and M.I.; visualization, I.S.; supervision, I.S., C.A.A.-M. and M.I.; project administration, I.S. and C.A.-M.; funding acquisition, I.S. and C.A.-M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Project Dicyt-062117SG, FONDEF No. ID21I10191, FONDECYT No. 1211132, STIC-AmSud 22-STIC-01, and BECAS DE MASTER NACIONAL ANID No. 22220262 in Chile.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: To the support of the University of Santiago's Postgraduate Vicerectory and Master's Program in Engineering Sciences in Electrical Engineering.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

2D	Two dimensional
3D	Three dimensional
AI	Artificial Intelligence
BER	Bit Error Rate
CIE	Comission Internationale de l'Éclairage
COVID-19	Coronavirus disease 2019
CSK	Colour Shift Keying
CSK/QAM	Colour shift keying with quadrature Amplitude modulation
CxR	Chest X-rays
DC	Direct Current
DNA	Deoxyribonucleic acid
DNN	Deep Neural Network
dsDNA	Double Strand DNA
ETC	Extra Trees Classifier
FSK	Frequency Shift Keying
GaussianNB	Naive Bayesian with Gaussian optimisation
GP	Gaussian Processes
IM/DD	Intensity-modulation direct-detection
LED	Light Emitting Diode
LGB	Light Gradient Boosting
LoS	Line-of-sight
LS	Light source
M	Modulation order
ML	Machine Learning
MMSE	Minimum mean square error
NLoS	Non-Line-of-sight
N^{DNA}	Chromosome length

OOK	On-off keying
OWC	Optical Wireless Communication
PI	Probability of improvement
PCR	Polymerase Chain Reaction
PDs	Photo detector
QAM	Quadrature amplitude modulation
RF	Radio Frequency
RGB	Red Green Blue
RNA	Ribonucleic acid
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
SNR	Signal-to-noise ratio
SPDs	Semiconductor Photo Detectors
SVM	Support vector machine
UM-VLC	Underground Mining Visible light communication
UM-VLC SISO	UM-VLC Single input single output
VLC	Visible light communication
VLC/FSK	Visible light communication/Frequency Shift Keying
VLS	Virtual light source
WBD	Water-borne Dataset

Appendix A

Reference LoS

$$H_{\text{LoS}} = \mathcal{C}\mathcal{G} \frac{(m+1)A_R}{2\pi d^2} \Omega_{\text{LoS}} \Psi_{\text{LoS}},$$

$$\mathcal{C} = \begin{cases} \frac{\eta}{\sin^2(\theta_0)} & 0 \leq \theta_0 \leq \Theta_{\text{FoV}} \\ 0 & \theta_0 > \Theta_{\text{FoV}}, \end{cases}$$

Reference NLoS

$$H_{\text{NLoS},(1)} = \frac{(m+1)A_R}{2\pi} \sum_{w=1}^W \mathcal{C}_w \mathcal{G}_w \frac{A_{\text{refl},w} \rho_w}{d_{1,w}^2 d_{2,w}^2} \Omega_{\text{NLoS}}^{(w)} \Psi_{\text{NLoS}}^{(w)},$$

Reference Scattering Model

$$H_{\text{scatter}} = \lim_{N \rightarrow \infty} \frac{(m+1)A_R \rho_s}{2N\pi} \sum_{n=1}^N \frac{p_n}{d_n^2} \Omega_{\text{scatter}}^{(n)} \Pi(\theta_n),$$

$$\Pi(\theta_n) = \begin{cases} 1 & 0 \leq \theta_n \leq \Theta_{\text{FoV}} \\ 0 & \theta_n > \Theta_{\text{FoV}}. \end{cases}$$

$$p_n = \frac{k_s^{\text{mie}}}{k_s} p_{\text{mie}}(\phi_n) + \frac{k_s^{\text{ray}}}{k_s} p_{\text{ray}}(\phi_n)$$

where: m is the Lambertian mode number,

A_R is the PD surface area,

d is the distance between LS and PD,

$\mathcal{G} \in [0, 1]$ is the optical filter gain,

$\mathcal{C} \in [1, +\infty)$ is the optical concentrator gain,

$\Omega_{\text{LoS}} \in [0, 1]$ is a trigonometric expression that depends on the relative position, rotation and tilt of the LS and PD, the half-power angle of the light beam, $\phi_{1/2}$, the Lambertian mode number, m , and the field of vision angle of the PD,

Θ_{FoV} ; and $\Psi_{\text{LoS}} \in [0, 1]$ is a scaling factor which accounts for the degree of shadowing in the LoS link

W is the number of reflectors, modelled as the generalized Lambertian virtual LSs, such that for the w -th reflector,

$A_{\text{refl},w}$ is its surface area,

$\rho_w \in [0, 1]$ is its reflectance,

$d_{1,w}$ is the distance from the LS to the reflector

$d_{2,w}$ is the distance from the reflector to the PD

\mathcal{G}_w and \mathcal{C}_w are the optical filter and the concentrator gain at the receiver, respectively, accounting for the incident angle from the w -th reflector to the PD,

$\Omega_{\text{NLoS}}^{(w)} \in [0, 1]$ is a trigonometric expression which depends on the relative position, rotation and tilt of the LS, the w -th reflector and the PD

$\Psi_{\text{NLoS}}^{(w)} \in [0, 1]$ is a scaling factor that accounts for the degree of shadowing in the w -th NLoS link. k_s^{mie} is the scattering coefficient of the Mie scattering dust particles,

k_s^{ray} is the scattering coefficient of the Rayleigh scattering air molecules,

$k_s = k_s^{\text{mie}} + k_s^{\text{ray}}$ is the total scattering coefficient,

$p_{\text{mie}}(\cdot)$ is the Mie scattering phase function,

$p_{\text{ray}}(\cdot)$ is the Rayleigh scattering phase function.

Appendix B

Next, we briefly explain the method used for constructing an extension field $GF(p^n)$, with $n \in \mathbb{Z}$ and $n > 1$, using $GF(p)$ as the underlying field, which was the subject of our research [29,30]. First, an irreducible polynomial $Q(x)$ of degree $n \in \mathbb{Z}, n > 1$ over $GF(p)$ is selected,

$$Q(x) = x^n + q_{n-1}x^{n-1} + \dots + q_1x + q_0,$$

where $q_i \in GF(p)$ for $i = 0, 1, \dots, n - 1$. The polynomial $Q(x)$ is called a primitive polynomial. Let α be a root of $Q(x)$, that is, $Q(\alpha) = 0$. Then,

$$\alpha^n = q'_{n-1}\alpha^{n-1} + \dots + q'_1\alpha + q'_0, \quad (\text{A1})$$

where q'_i is the additive inverse of q_i [29]. Next, α^{n+1} is constructed recursively as

$$\begin{aligned} \alpha^{n+1} &= \alpha^* \alpha^n = \alpha^* [q'_{n-1}\alpha^{n-1} + \dots + q'_1\alpha + q'_0] \\ \alpha^{n+1} &= q'_{n-1}\alpha^n + \dots + q'_1\alpha^2 + q'_0\alpha, \end{aligned}$$

and the element α^n is replaced using Equation (A1),

$$\alpha^{n+1} = q'_{n-1} [q'_{n-1}\alpha^{n-1} + \dots + q'_1\alpha + q'_0] + \dots + q'_1\alpha^2 + q'_0\alpha,$$

then:

$$\alpha^{n+1} = a_{n-1}\alpha^{n-1} + \dots + a_1\alpha + a_0,$$

where $a_{n-1} = q'_{n-1} * q'_{n-1}, \dots, a_1 = q'_{n-1} * q'_1 + q'_0$ and $a_0 = q'_{n-1} * q'_0$. Thus, the nonzero elements of $GF(p^n)$ are generated as linear combinations of $\{1, \alpha, \alpha^2, \dots, \alpha^{n-1}\}$ in the following manner:

$$\alpha^k = a_{n-1}\alpha^{n-1} + \dots + a_1\alpha + a_0, \quad (\text{A2})$$

with $k \geq 0, a_i \in GF(p), i = 0, 1, \dots, n - 1$.

References

1. Dhama, K.; Chandran, D.; Chopra, H.; Islam, M.A.; Emran, T.B.; Rehman, M.E.U.; Dey, A.; Mohapatra, R.K.; SV, P.; Mohankumar, P.; et al. SARS-CoV-2 emerging Omicron subvariants with a special focus on BF.7 and XBB.1.5 recently posing fears of rising cases amid ongoing COVID-19 pandemic. *J. Exp. Biol. Agric. Sci.* **2022**, *10*, 1215–1221. [[CrossRef](#)]
2. Carvajal, J.; Casanello, P.; Toso, A.; Fariás, M.; Carrasco-Negue, K.; Araujo, K.; Valero, P.; Fuenzalida, J.; Solari, C.; Sobrevia, L. Functional consequences of SARS-CoV-2 infection in pregnant women, fetoplacental unit, and neonate. *Biochim. Biophys. Acta (BBA) Mol. Basis Dis.* **2023**, *1869*, 166582. [[CrossRef](#)] [[PubMed](#)]
3. Coppée, R.; Chenane, H.R.; Bridier-Nahmias, A.; Tcherakian, C.; Catherinot, E.; Collin, G.; Lebourgeois, S.; Visseaux, B.; Descamps, D.; Vasse, M.; et al. Temporal dynamics of RSV shedding and genetic diversity in adults during the COVID-19 pandemic in a French hospital, early 2021. *Virus Res.* **2023**, *323*, 198950. [[CrossRef](#)] [[PubMed](#)]
4. Khalid, M.; Murphy, D.; Shoai, M.; George-William, J.N.; Al-ebini, Y. Geographical distribution of host's specific SARS-CoV-2 mutations in the early phase of the COVID-19 pandemic. *Gene* **2023**, *851*, 147020. [[CrossRef](#)]
5. Dhama, K.; Nainu, F.; Frediansyah, A.; Yatoo, M.I.; Mohapatra, R.K.; Chakraborty, S.; Zhou, H.; Islam, M.R.; Mamada, S.S.; Kusuma, H.I.; et al. Global emerging Omicron variant of SARS-CoV-2: Impacts, challenges and strategies. *J. Infect. Public Health* **2023**, *16*, 4–14. [[CrossRef](#)]
6. Alsaedy, A.A.R.; Chong, E.K.P. Detecting Regions At Risk for Spreading COVID-19 Using Existing Cellular Wireless Network Functionalities. *IEEE Open J. Eng. Med. Biol.* **2020**, *1*, 187–189. [[CrossRef](#)]
7. Adams, S.D.; Valentine, A.; Bucknall, T.K.; Kouzani, A.Z. Technologies for Fever Screening in the Time of COVID-19: A Review. *IEEE Sens. J.* **2022**, *22*, 16720–16729. [[CrossRef](#)]
8. Gad, A.; ElBary, G.; Alkhedher, M.; Ghazal, M. Vision-based Approach for Automated Social Distance Violators Detection. In Proceedings of the 2020 International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT), Sakheer, Bahrain, 20–21 December 2020; pp. 1–5. [[CrossRef](#)]
9. Quilliam, R.S.; Weidmann, M.; Moresco, V.; Purshouse, H.; O'Hara, Z.; Oliver, D.M. COVID-19: The environmental implications of shedding SARS-CoV-2 in human faeces. *Environ. Int.* **2020**, *140*, 105790. [[CrossRef](#)]
10. Ahmed, W.; Bivins, A.; Simpson, S.L.; Bertsch, P.M.; Ehret, J.; Hosegood, I.; Metcalfe, S.S.; Smith, W.J.; Thomas, K.V.; Tynan, J.; et al. Wastewater surveillance demonstrates high predictive value for COVID-19 infection on board repatriation flights to Australia. *Environ. Int.* **2022**, *158*, 106938. [[CrossRef](#)]
11. Murphy, H.M.; Thomas, M.K.; Schimidt, P.J.; Medeiros, D.T.; McFADYEN, S.; PINTAR, K.D.M. Estimating the burden of acute gastrointestinal illness due to Giardia, Cryptosporidium, Campylobacter, E. coli O157 and norovirus associated with private wells and small water systems in Canada. *Epidemiol. Infect.* **2016**, *144*, 1355–1370. [[CrossRef](#)]
12. Wright, C.J.; Sargeant, J.M.; Edge, V.L.; Ford, J.D.; Farahbakhsh, K.; Shiwak, I.; Flowers, C.; Harper, S.L. Water quality and health in northern Canada: stored drinking water and acute gastrointestinal illness in Labrador Inuit. *Environ. Sci. Pollut. Res.* **2018**, *25*, 32975–32987. [[CrossRef](#)] [[PubMed](#)]
13. Abdulkadir, N.; O. Afolabi, R.; M. Usman, H.; Mustapha, G.; A. Abubakar, U. Epidemiological Studies of Waterborne Diseases in Relation to Bacteriological Quality of Water. *Microbiol. Res. J. Int.* **2019**, *28*, 1–12. [[CrossRef](#)]
14. Tiwari, S.B.; Gahlot, P.; Tyagi, V.K.; Zhang, L.; Zhou, Y.; Kazmi, A.; Kumar, M. Surveillance of Wastewater for Early Epidemic Prediction (SWEEP): Environmental and health security perspectives in the post COVID-19 Anthropocene. *Environ. Res.* **2021**, *195*, 110831. [[CrossRef](#)] [[PubMed](#)]
15. Thevenot, J.; Lopez, M.B.; Hadid, A. A Survey on Computer Vision for Assistive Medical Diagnosis From Faces. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1497–1511. [[CrossRef](#)] [[PubMed](#)]
16. Pathak, N.; Deb, P.K.; Mukherjee, A.; Misra, S. IoT-to-the-Rescue: A Survey of IoT Solutions for COVID-19-Like Pandemics. *IEEE Internet Things J.* **2021**, *8*, 13145–13164. [[CrossRef](#)]
17. Mary, L.W.; Raj, S.A.A. A Survey on SARS-COV-2 (COVID-19) using Machine Learning Techniques. In Proceedings of the 2021 2nd International Conference on Smart Electronics and Communication (ICOSEC), Trichy, India, 7–9 October 2021; pp. 1612–1617. [[CrossRef](#)]
18. Shen, Y.; Guo, D.; Long, F.; Mateos, L.A.; Ding, H.; Xiu, Z.; Hellman, R.B.; King, A.; Chen, S.; Zhang, C.; et al. Robots Under COVID-19 Pandemic: A Comprehensive Survey. *IEEE Access* **2021**, *9*, 1590–1615. [[CrossRef](#)]
19. Leung, C.K.; Chen, Y.; Shang, S.; Deng, D. Big Data Science on COVID-19 Data. In Proceedings of the 2020 IEEE 14th International Conference on Big Data Science and Engineering (BigDataSE), Guangzhou, China, 31 December 2020–1 January 2021; pp. 14–21. [[CrossRef](#)]
20. Silveira, T.M.; Pinho, P.; Carvalho, N.B. RFID Tattoo for COVID-19 Temperature Measuring. In Proceedings of the 2021 IEEE Radio and Wireless Symposium (RWS), San Diego, CA, USA, 17–22 January 2021; pp. 98–100. [[CrossRef](#)]
21. Lubecke, L.C.; Ishmael, K.; Zheng, Y.; Boric-Lubecke, O.; Lubecke, V.M. Identification of COVID-19 Type Respiratory Disorders Using Channel State Analysis of Wireless Communications Links. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico, 1–5 November 2021; pp. 7582–7585. [[CrossRef](#)]
22. Perumal, V.; Theivanithy, K. A Transfer Learning Model for COVID-19 Detection with Computed Tomography and Sonogram Images. In Proceedings of the 2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 25–27 March 2021; pp. 80–83. [[CrossRef](#)]

23. Fernandez–Grandon, C.; Soto, I.; Zabala-Blanco, D.; Alavia, W.; Garcia, V. SVM and ANN classification using GLCM and HOG features for COVID-19 and Pneumonia detection from Chest X-rays. In Proceedings of the 2021 Third South American Colloquium on Visible Light Communications (SACVLC), Toledo, Brazil, 11–12 November 2021; pp. 1–6. [[CrossRef](#)]
24. Jativa, P.P.; Azurdia-Meza, C.A.; Canizares, M.R.; Cespedes, S.; Montejo-Sanchez, S. Performance Enhancement of VLC-Based Systems Using Diversity Combining Schemes in the Receiver. In Proceedings of the 2019 IEEE Latin-American Conference on Communications (LATINCOM), Salvador, Brazil, 11–13 November 2019; pp. 1–6. [[CrossRef](#)]
25. Palacios Játiva, P.; Román Cañizares, M.; Azurdia-Meza, C.A.; Zabala-Blanco, D.; Dehghan Firoozabadi, A.; Seguel, F.; Montejo-Sánchez, S.; Soto, I. Interference Mitigation for Visible Light Communications in Underground Mines Using Angle Diversity Receivers. *Sensors* **2020**, *20*, 367. [[CrossRef](#)]
26. Román Cañizares, M.; Palacios Játiva, P.; Azurdia-Meza, C.A.; Montejo-Sánchez, S.; Céspedes, S. Impact of diversity combining schemes in a multi-cell VLC system with angle diversity receivers. *Photonic Netw. Commun.* **2022**, *43*. [[CrossRef](#)]
27. Seguel, F.; Palacios-Jativa, P.; Azurdia-Meza, C.A.; Krommenacker, N.; Charpentier, P.; Soto, I. Underground Mine Positioning: A Review. *IEEE Sensors J.* **2022**, *22*, 4755–4771. [[CrossRef](#)]
28. Soto, I.; Nilson Rodrigues, R.; Massuyama, G.; Seguel, F.; Palacios Játiva, P.; Azurdia-Meza, C.A.; Krommenacker, N. A Hybrid VLC-RF Portable Phasor Measurement Unit for Deep Tunnels. *Sensors* **2020**, *20*, 790. [[CrossRef](#)]
29. Jirón, I.; Soto, S.; Marín, S.; Acosta, M.; Soto, I. A new DNA-based model for finite field arithmetic. *Heliyon* **2019**, *5*, e02901. [[CrossRef](#)] [[PubMed](#)]
30. Jiron, I.; Soto, I.; Azurdia-Meza, C.A.; Valencia, A.; Carrasco, R. A new DNA cryptosystem based on AG codes evaluated in gaussian channels. *Telecommun. Syst.* **2017**, *64*, 279–291. [[CrossRef](#)]
31. Soto, I.; Jiron, I.; Valencia, A.; Carrasco, R. Secure DNA data compression using algebraic curves. *Electron. Lett.* **2015**, *51*, 1466–1468. [[CrossRef](#)]
32. Zamorano-Illanes, R.; Estela, M.C.; Soto, I.; Ijaz, M.; Rau, F. MIMO QAM indoor VLC using polar codes for low-cost emitters and FPGA receiver. In Proceedings of the 2022 4th West Asian Symposium on Optical and Millimeter-wave Wireless Communications (WASOWC), Tabriz, Iran, Islamic Republic, 12–13 May 2022; pp. 1–7. [[CrossRef](#)]
33. Ayub, M.S.; Wuttisittikulkij, L.; Adasme, P.; Soto, I. Hybrid Precoding Design for Two Carriers Aggregated in 5G Massive MIMO System. In Proceedings of the 2020 South American Colloquium on Visible Light Communications (SACVC), Santiago, Chile, 4–5 June 2020; pp. 1–5. [[CrossRef](#)]
34. Adasme, P.; Soto, I.; Juan, E.S.; Seguel, F.; Firoozabadi, A.D. Maximizing Signal to Interference Noise Ratio for Massive MIMO: A Mathematical Programming Approach. In Proceedings of the 2020 South American Colloquium on Visible Light Communications (SACVC), Santiago, Chile, 4–5 June 2020; pp. 1–6. [[CrossRef](#)]
35. Duncan, A.; Janssen, M. Pascual Jordan’s resolution of the conundrum of the wave-particle duality of light. *Stud. Hist. Philos. Sci. Part B Stud. Hist. Philos. Mod. Phys.* **2008**, *39*, 634–666. [[CrossRef](#)]
36. Rahaim, M.B.; Little, T.D. Toward practical integration of dual-use VLC within 5G networks. *IEEE Wirel. Commun.* **2015**, *22*. [[CrossRef](#)]
37. Wang, J.; Al-Kinani, A.; Zhang, W.; Wang, C.X.; Zhou, L. A general channel model for visible light communications in underground mines. *China Commun.* **2018**, *15*, 95–105. [[CrossRef](#)]
38. Jativa, P.P.; Azurdia-Meza, C.A.; Sanchez, I.; Seguel, F.; Zabala-Blanco, D.; Firoozabadi, A.D.; Gutierrez, C.A.; Soto, I. A VLC Channel Model for Underground Mining Environments With Scattering and Shadowing. *IEEE Access* **2020**, *8*, 185445–185464. [[CrossRef](#)]
39. Sun, Y.; Gong, C.; Xu, Z.; Zhan, Y. Link Gain and Pulse Width Broadening Evaluation of Non-Line-of-Sight Optical Wireless Scattering Communication Over Broad Spectra. *IEEE Photonics J.* **2017**, *9*, 1–12. [[CrossRef](#)]
40. Liu, W.; Zou, D.; Xu, Z. Modeling of optical wireless scattering communication channels over broad spectra. *J. Opt. Soc. Am. A* **2015**, *32*, 486. [[CrossRef](#)]
41. Kong, S.H. TOA and AOD statistics for down link Gaussian scatterer distribution model. *IEEE Trans. Wirel. Commun.* **2009**, *8*, 2609–2617. [[CrossRef](#)]
42. Borhani, A.; Patzold, M. Time-of-arrival, angle-of-arrival, and angle-of-departure statistics of a novel simplistic disk channel model. In Proceedings of the 2011 5th International Conference on Signal Processing and Communication Systems (ICSPCS), Honolulu, HI, USA, 12–14 December 2011; pp. 1–7. [[CrossRef](#)]
43. Borhani, A.; Patzold, M. A Unified Disk Scattering Model and Its Angle-of-Departure and Time-of-Arrival Statistics. *IEEE Trans. Veh. Technol.* **2013**, *62*, 473–485. [[CrossRef](#)]
44. Tennakoon, P.; Wavegedara, C.B. A GBSM Indoor Channel Model With an Arbitrary Center Point of Gaussian Scatterer Distribution. *IEEE Trans. Antennas Propag.* **2022**, *70*, 2128–2136. [[CrossRef](#)]
45. Harish Kalla, L.M. Design of Optical Light Communication System: Study The Effect of Light Wavelength on Transmission Efficiency Using Audio Signals. *Int. J. Eng. Sci. Res. Technol.* **2016**, *5*, 622–627. [[CrossRef](#)]
46. Chow, C.W.; Shiu, R.J.; Liu, Y.C.; Liao, X.L.; Lin, K.H.; Wang, Y.C.; Chen, Y.Y. Using advertisement light-panel and CMOS image sensor with frequency-shift-keying for visible light communication. *Opt. Express* **2018**, *26*. [[CrossRef](#)]
47. Salmento, M.L.G.; Soares, G.M.; Alonso, J.M.; Braga, H.A. A dimmable offline LED driver with OOK-M-FSK modulation for VLC applications. *IEEE Trans. Ind. Electron.* **2019**, *66*. [[CrossRef](#)]

48. Dahri, F.A.; Mangrio, H.B.; Baqai, A.; Umrani, F.A. Experimental Evaluation of Intelligent Transport System with VLC Vehicle-to-Vehicle Communication. *Wirel. Pers. Commun.* **2019**, *106*. [[CrossRef](#)]
49. Guajardo-Penroz, C.; Soto, I.; San-Juan, E.; Adasme, P.; Azurdia-Meza, C.; Zabala-Blanco, D. Finite Field Metrics Applied to M-FSK Modulation in VLC Systems. In Proceedings of the 2020 South American Colloquium on Visible Light Communications (SACVC), Santiago, Chile, 4–5 June 2020; pp. 1–6. [[CrossRef](#)]
50. Guajardo-Penroz, C.; Soto, I.; San-Juan, E.; Adasme, P.; Azurdia-Meza, C.; Alavia, W. New High Dimming Range M-FSK Demodulation Strategy for VLC Systems. In Proceedings of the 2020 12th International Symposium on Communication Systems, Networks and Digital Signal Processing (CSNDSP), Porto, Portugal, 20–23 July 2020; pp. 1–6. [[CrossRef](#)]
51. Madrid, E.D.; Rojas Lobos, F.; Soto, I.; Gatica, G. Hysteresis based FSK modulation for visible light communication. In Proceedings of the 2020 South American Colloquium on Visible Light Communications (SACVC), Santiago, Chile, 4–5 June 2020; pp. 1–6. [[CrossRef](#)]
52. Morris, A. The Dual Nature of Light. In *Why Icebergs Float*; UCL Press: London, UK, 2016; pp. 31–49. [[CrossRef](#)]
53. Aharonov, Y.; Cohen, E.; Colombo, F.; Landsberger, T.; Sabadini, I.; Struppa, D.C.; Tollaksen, J. Finally making sense of the double-slit experiment. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 6480–6485. [[CrossRef](#)]
54. Pepe, A.; Wei, Z.; Fu, H.Y. Heuristic, machine learning approach to 8-CSK decision regions in RGB-LED visible light communication. *OSA Contin.* **2020**, *3*, 473. [[CrossRef](#)]
55. Zhang, D.F.; Yu, H.Y.; Zhu, Y.J. A multi-user joint constellation design of color-shift keying for VLC downlink broadcast channels. *Opt. Commun.* **2020**, *473*, 126001. [[CrossRef](#)]
56. Dong, Z.; Yu, J.; Chen, Y.; Li, F.; Xin, X. Symbol division multiplexing in optical fiber communication systems. *Opt. Express* **2022**, *30*, 14998. [[CrossRef](#)]
57. Zhang, Y.; Wang, H.; Ji, Y.; Zhang, Y. Parallel Distribution Matcher Base on CCDM for Probabilistic Amplitude Shaping in Coherent Optical Fiber Communication. *Photonics* **2022**, *9*, 604. [[CrossRef](#)]
58. Hart, J. Terraforming mars and marsforming terra: Discovery doctrine in space. *Theol. Sci.* **2019**, *17*. [[CrossRef](#)]
59. Babukov, Y.; Aleksandrov, R.; Ivanova, A.; Atemin, A.; Stoynov, S. DNAREpairK: An Interactive Database for Exploring the Impact of Anticancer Drugs onto the Dynamics of DNA Repair Proteins. *Biomedicines* **2021**, *9*, 1238. [[CrossRef](#)]
60. Touati, R.; Messaoudi, I.; Oueslati, A.; Lachiri, Z.; Kharrat, M. New Intra-class Helitrons Classification Using DNA-Image Sequences and Machine Learning Approaches. *IRBM* **2021**, *42*, 154–164. [[CrossRef](#)]
61. Guajardo, J.; Güneysu, T.; Kumar, S.S.; Paar, C.; Pelzl, J. Efficient Hardware Implementation of Finite Fields with Applications to Cryptography. *Acta Appl. Math.* **2006**, *93*, 75–118. [[CrossRef](#)]
62. Cusick, T.W.; Koblitz, N. Algebraic Aspects of Cryptography. *Am. Math. Mon.* **2000**, *107*, 384. [[CrossRef](#)]
63. Nawaz, S.J.; Sharma, S.K.; Wyne, S.; Patwary, M.N.; Asaduzzaman, M. Quantum Machine Learning for 6G Communication Networks: State-of-the-Art and Vision for the Future. *IEEE Access* **2019**, *7*, 46317–46350. [[CrossRef](#)]
64. Yu, T.C.; Huang, W.T.; Lee, W.B.; Chow, C.W.; Chang, S.W.; Kuo, H.C. Visible Light Communication System Technology Review: Devices, Architectures, and Applications. *Crystals* **2021**, *11*, 1098. [[CrossRef](#)]
65. Ai, T.; Yang, Z.; Hou, H.; Zhan, C.; Chen, C.; Lv, W.; Tao, Q.; Sun, Z.; Xia, L. Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases. *Radiology* **2020**, *296*, E32–E40. [[CrossRef](#)]
66. Ardila, D.; Kiraly, A.P.; Bharadwaj, S.; Choi, B.; Reicher, J.J.; Peng, L.; Tse, D.; Etemadi, M.; Ye, W.; Corrado, G.; et al. Author Correction: End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat. Med.* **2019**, *25*, 1319–1319. [[CrossRef](#)]
67. Krishnamachari, K.; Lu, D.; Swift-Scott, A.; Yeraliyev, A.; Lee, K.; Huang, W.; Leng, S.N.; Skanderup, A.J. Accurate somatic variant detection using weakly supervised deep learning. *Nat. Commun.* **2022**, *13*, 4248. [[CrossRef](#)] [[PubMed](#)]
68. Figueroa, S.; Freire-Paspuel, B.; Vega-Mariño, P.; Velez, A.; Cruz, M.; Cardenas, W.B.; Garcia-Bereguian, M.A. High sensitivity-low cost detection of SARS-CoV-2 by two steps end point RT-PCR with agarose gel electrophoresis visualization. *Sci. Rep.* **2021**, *11*, 21658. [[CrossRef](#)] [[PubMed](#)]
69. Ortiz, S.; Rojas, F.; Valenzuela, O.; Herrera, L.J.; Rojas, I. Determination of the Severity and Percentage of COVID-19 Infection through a Hierarchical Deep Learning System. *J. Pers. Med.* **2022**, *12*, 535. [[CrossRef](#)] [[PubMed](#)]
70. Mansoor, R.; Shah, M.A.; Khattak, H.A.; Mussadiq, S.; Rauf, H.T.; Ameer, Z. Detection of Diseases in Pandemic: A Predictive Approach Using Stack Ensembling on Multi-Modal Imaging Data. *Electronics* **2022**, *11*, 3974. [[CrossRef](#)]
71. Hoehner, P.A. *Visible Light Communications: Theoretical and Practical Foundations*; Carl Hanser Verlag GmbH Co KG: Munich, Germany, 2019.
72. Niu, W.; Xu, Z.; Xiao, W.; Liu, Y.; Hu, F.; Wang, G.; Zhang, J.; He, Z.; Yu, S.; Shi, J.; et al. Phosphor-Free Golden Light LED Array for 5.4-Gbps Visible Light Communication Using MIMO Tomlinson-Harashima Precoding. *J. Light. Technol.* **2022**, *40*, 5031–5040. [[CrossRef](#)]
73. Van de Beek, J.J.; Edfors, O.; Sandell, M.; Wilson, S.; Borjesson, P. On channel estimation in OFDM systems. In Proceedings of the 1995 IEEE 45th Vehicular Technology Conference. Countdown to the Wireless Twenty-First Century, Chicago, IL, USA, 25–28 July 1995; Volume 2, pp. 815–819. [[CrossRef](#)]

74. IEEE Computer Society; LAN/MAN Standards Committee; Institute of Electrical and Electronics Engineers; IEEE-SA Standards Board. *IEEE Standard for Local and Metropolitan Area Networks. Part 15.7, Short-Range Wireless Optical Communication Using Visible Light*; Institute of Electrical and Electronics Engineers: Piscataway Township, NJ, USA 2011; p. 286. [[CrossRef](#)]
75. Géron, A. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 2nd ed.; O'Reilly: Sebastopol, CA, USA, 2019.
76. Suykens, J.A.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [[CrossRef](#)]
77. Jiang, Q.; Huang, B.; Yan, X. GMM and optimal principal components-based Bayesian method for multimode fault diagnosis. *Comput. Chem. Eng.* **2016**, *84*. [[CrossRef](#)]
78. Kushner, H.J. A New Method of Locating the Maximum Point of an Arbitrary Multippeak Curve in the Presence of Noise. *J. Basic Eng.* **1964**, *86*, 97–106. [[CrossRef](#)]
79. Srinivas, N.; Krause, A.; Kakade, S.; Seeger, M. Gaussian process optimization in the bandit setting: No regret and experimental design. In Proceedings of the ICML 2010—Proceedings, 27th International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010. [[CrossRef](#)]
80. Goldberg, Y.; Elhadad, M. splitSVM: Fast, Space-efficient, non-Heuristic, polynomial kernel computation for NLP applications. In Proceedings of the ACL-08: HLT—46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference, Columbus, OH, USA, 6–17 June 2008.
81. Hsieh, C.J.; Chang, K.W.; Lin, C.J.; Keerthi, S.S.; Sundararajan, S. A dual coordinate descent method for large-scale linear SVM. In Proceedings of the 25th international conference on Machine learning—ICML '08, Helsinki Finland, 5–9 June 2008; ACM Press: New York, NY, USA, 2008; pp. 408–415. [[CrossRef](#)]
82. Tanha, J.; van Someren, M.; Afsarmanesh, H. Semi-supervised self-training for decision tree classifiers. *Int. J. Mach. Learn. Cybern.* **2017**, *8*, 355–370. [[CrossRef](#)]
83. Geurts, P.; Ernst, D.; Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **2006**, *63*, 3–42. [[CrossRef](#)]
84. Zhao, G. A New Perspective on Classification. Ph.D. Thesis, Utah State University, Logan, UT, USA, 2000. [[CrossRef](#)]
85. Padhi, D.K.; Padhy, N.; Bhoi, A.K.; Shafi, J.; Ijaz, M.F. A fusion framework for forecasting financial market direction using enhanced ensemble models and technical indicators. *Mathematics* **2021**, *9*, 2646. [[CrossRef](#)]
86. Chen, Y.; Xiong, J.; Xu, W.; Zuo, J. A novel online incremental and decremental learning algorithm based on variable support vector machine. *Clust. Comput.* **2019**, *22*, 7435–7445. [[CrossRef](#)]
87. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
88. Chen, T.; Guestrin, C. XGBoost. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; pp. 785–794. [[CrossRef](#)]
89. Banerjee, A.; Chitnis, U.; Jadhav, S.; Bhawalkar, J.; Chaudhury, S. Hypothesis testing, type I and type II errors. *Ind. Psychiatry J.* **2009**, *18*, 127. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.