

AN IOT-BASED COMPUTATIONAL INTELLIGENCE MODEL TO PERFORM GENE ANALYTICS IN PATERNITY TESTING AND COMPARISON FOR HEALTH 4.0

VIJAY ARPUTHARAJ J¹, K. SANKAR², A. SENTHIL KUMAR³, MULAGUNDLA SRIDEVI⁴,
D.DURGA PRASAD⁵

¹Assistant Professor, Department of Computer Science, CHRIST (Deemed to be University), India

²Associate professor, Department of Computer Science and Engineering (Data Science), CVR College of Engineering, Telangana, Hyderabad, India

³Dean & Associate Professor, Department of Computer Science, Skyline University, Nigeria.

⁴Associate Professor, Department of Computer Science and Engineering, CVR College of Engineering, Mangalpalli (V), Ibrahimpatnam (M), R R District, Telangana – 501 510

⁵Assistant Professor, Department of ECE, Vishnu Institute of Technology, Bhimavaram, AP, India

E-mail: ¹phd@gmail.com, ²sankarkrish@cvr.ac.in, ³senthask@gmail.com,

⁴m.sridevi@cvr.ac.in, sreetech99@gmail.com, ⁵durgaprasad.d@vishnu.edu.in

ABSTRACT

Parental comparison and parenthood testing are essential in various legal and medical scenarios. The accuracy and reliability of these tests heavily rely on the gene analysis algorithms used. However, analyzing the quality of succession data are quite challenging due to the presence of detrimental characteristics. To address this issue, we propose using machine learning-based algorithms such as clustering (Correlation-based) and Classification (Modified Naive Bayesian) to separate these characteristics from the parent-child gene array. This progression helps to identify, validate, and select tools, techniques for scrutinizing indecent sequences, leading to accurate and reliable results. In this paper, we present an IoT-based intelligence tool for parental comparison that uses a secure gene analysis algorithm. The model employs multiple sensors and devices to collect genetic data, which is then securely processed and analyzed using contemporary algorithms. The suggested model uses advanced techniques such as encryption and decryption to ensure the privacy and confidentiality of the genetic information. Our experimental consequences reveal that the proposed model is reliable, secure, and provides accurate results. The model has the potential to be used in various legal and medical contexts where the security and reliability of genetic data are critical.

Keywords: *Iot, Correlation Clustering, Gene Analysis, Paternity Testing.*

1. INTRODUCTION

In the era of Health 4.0, smart devices play a crucial role in learning specific tasks through the use of a recent advancements called CI - Computational Intelligence, which includes Fuzzy logic, Neural networks, Evolutionary computation, and fuzzy computation based on Learning theory, logic, Probabilistic methods [1][2][3]. These Computational Intelligence methods provide substantial benefits to the Internet of Things (IoT) intellectual devices, enabling the collection, processing, and study of massive amounts of health-related data in real-time.

One of the most significant advancements in Health 4.0 is in genomic studies, particularly in DNA gene mining and sequencing. With several

variants and sorts of genomic data and gene classifications available from various sources, these genomic sequences are now applied and used not only for research studies but also in personalized medicine. The integration of genomics data with other health data sources using IoT devices and big data analytics provides a new paradigm in healthcare, enabling precision medicine approaches for diagnosis, treatment, and disease prevention.

Health 4.0 is a term used to describe the fourth industrial revolution in healthcare, where digital technologies and data analytics are transforming the way healthcare is delivered, managed, and experienced. It represents a shift from a reactive, disease-focused healthcare model to a proactive, personalized, and patient-centered approach that

emphasizes prevention, early detection, and data-driven decision-making.

IoT-based DNA sequence analysis involves the use of various tools and techniques to collect, process, and analyze genetic data. Some of these tools and techniques include IoT Sensors, Cloud computing, Big Data Analytics, Data visualization etc. Overall, these tools and techniques enable scientists to collect and analyze genetic data efficiently, leading to better diagnosis and treatment decisions in healthcare.

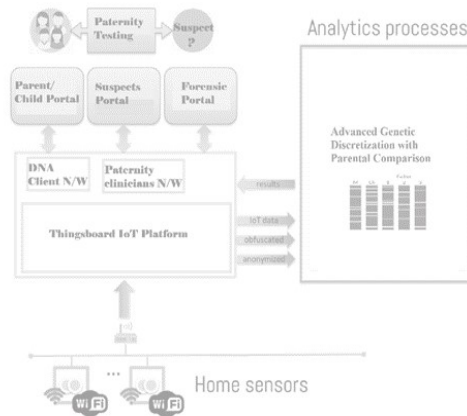


Figure 1.1. Architecture of IoT-based Computational Intelligence Model

Some of these projects demonstrate the potential of IoT-based gene analysis in advancing genomic medicine and improving healthcare outcomes in the era of Health 4.0.

In recent years, advancements in machine learning and artificial intelligence have further improved the accuracy and efficiency of IoT-based gene sequence analysis. These techniques enable researchers to identify patterns and insights in genetic data that would be difficult to detect manually, leading to better diagnosis and treatment decisions in healthcare.

IoT-based gene sequence analysis has played a significant role in advancing genomic medicine in the era of Health 4.0. The roots of IoT-based gene sequence analysis can be traced back to the Human Genome Project, which was completed in 2003. This project identified and mapped the genes in the human genome, laying the foundation for modern genomic medicine.

The advent of IoT technology and big data analytics in the early 2010s revolutionized gene sequence analysis, making it more accessible and affordable. IoT devices such as wearable health monitors, medical sensors, and genomic sequencers can now collect vast amounts of genetic data in

real-time. The use of cloud computing platforms and big data analytics tools enables researchers to analyze this genetic data quickly and efficiently.

The core interest area of the article "An IoT-Based Computational Intelligence Model to Perform Gene Analytics in Paternity Testing and Comparison for Health 4.0" revolves around the intersection of IoT, computational intelligence, gene analytics, paternity testing, and their application in the context of Health 4.0. Specifically, the paper aims to explore how an IoT-based computational intelligence model can be utilized to analyze genes for the purpose of paternity testing and comparison, within the framework of Health 4.0.

The integration of IoT-based gene sequence analysis with other health data sources has also led to the emergence of personalized medicine. By combining genomic data with other health data such as medical records, family history, and lifestyle factors, clinicians can provide personalized treatment plans tailored to an individual's unique genetic makeup.

Table 1.1 Advancements in IoT based Projects

Year	IoT based Projects & Advancements
2003 - Human Genome Project	This project identified and mapped all the genes in the human genome, laying the foundation for modern genomic medicine.
2005 - The Personal Genome Project	It aims to create an open-access genomic database by collecting and analyzing genetic data from thousands of volunteers using IoT devices such as wearable health monitors and genomic sequencers.
2006 - The Cancer Genome Atlas	It aims to identify the genomic alterations that underlie different types of cancer. The project collects and analyzes genomic data from thousands of cancer patients using IoT devices such as genomic sequencers and medical sensors.
2012 - The 100,000 Genomes Project	This project aims to sequence and analyze the genomes of 100,000 individuals with rare diseases and cancer using IoT devices and big data analytics. The project aims to improve diagnoses and treatment decisions for these conditions.
2013 - The Global Alliance	This international collaboration, launched in 2013, aims to develop standards and guidelines for the

for Genomics and Health	responsible sharing of genomic data. The alliance uses IoT devices and big data analytics to collect and analyze genomic data from around the world.
2015 - The Precision Medicine Initiative	This initiative aims to accelerate biomedical research and improve healthcare outcomes by integrating genomics and other health data sources using IoT devices and big data analytics.

and devices. The study will involve the collection of gene data from individuals and the development of a secure algorithm to analyze the data. The proposed algorithm will be tested using simulated and real gene data, and its performance will be compared with existing paternity testing methods. The research will be limited to gene data analysis and paternity testing using an IoT-based computational intelligence model in healthcare 4.0.

1.2. Research Objectives

The primary goal and objective of the current study is to develop an IoT based advanced gene analysis algorithm for paternity testing and parental comparison. The algorithm will use an IoT-based computational intelligence model in healthcare 4.0 to assure the efficiency, reliability, and correctness of the outcomes. The specific objectives of the research are:

- To design an IoT-based CI – Computational Intelligence model for genomic data analysis for hereditary disease identification, medical diagnosis, and parental comparisons.
- To develop a secure gene analysis algorithm for parental comparison and paternity testing based on IoT and related technologies.
- To evaluate the success of the current model in terms of efficiency, reliability, correctness, and security.
- To compare the generated results of the IoT based algorithm with existing paternity testing methods.
- To validate the proposed algorithm through experimental studies.

In summary, this study's significance lies in its contribution to the healthcare industry's progress towards healthcare 4.0. It presents a solution that can efficiently handle and analyze large amounts of genome sequence data while integrating IoT technology, which can lead to more effective disease diagnosis and treatment.

1.3. Scope

The proposed research focuses on developing a secure gene analysis algorithm for paternity testing and parental comparison. The algorithm will be designed using an IoT-based computational intelligence model, which will enable the integration of data from various sources

1.4. Limitation

The proposed research has several limitations, including:

- The availability and quality of gene data from individuals may vary, which can affect the correctness of the results.
- The proposed algorithm may be limited to specific types of genetic markers, which may not be suitable for all types of paternity testing.
- The research will be limited to the IoT-based computational intelligence model and may not consider other factors that may affect paternity testing.
- The study will be limited to experimental studies, and the results may not be generalized to all populations.
- The proposed algorithm may require high computational power, which may limit its scalability and practicality.

Table 2.1 Research Major Breakthroughs

Research Phase	Major Breakthrough
Phase-1: Sensor Automation	Successful activation of DNA sequencing sensors
Phase-2: Network and Protocols Automation	Network and Protocols - WSN, Zigbee, TLS, SSL, FASTQ, BAM, and VCF
Phase-3: Real time analytics	Analytics Process - Clustering & Classification, Pattern matching
Phase-4: IoT App inclusion & Management	IoT – Security, Data management, Data processing (Still in Progress)

2. LITERATURE REVIEW

The subsequent section contains some of the articles related to base tools and techniques in Health 4.0 (Both IoT based techniques and gene sequencing techniques).

2.1 Statement of problem

IoT-based paternity testing presents several challenges that require attention to provide accurate and reliable results. The inspiration for this

research stems from three primary areas of research advancement namely, biological and medical diagnosis, gene mining and data analysis, and machine learning algorithms.

The proposed gene-based advanced model contains a significant contribution to the healthcare industry. This study utilizes a hybrid approach for patient genome sequence analysis, which addresses the challenges of IoT-based paternity testing in Healthcare 4.0 with the futuristic view of Healthcare 5.0.

Parental Comparison and sequence matching is a crucial technique for paternity testing analysis and health analytics, as it enables the identification of disease outbreaks at a faster pace. However, the need for healthcare 4.0 demands more efficient systems that may professionally interact with large data. The integration of IoT technology in healthcare 4.0 can potentially provide smart gadgets that can address these requirements.

This study presents an IoT-enabled computational intelligence model that employs an advanced gene analysis algorithm for parental comparison and paternity testing. This model can handle large amounts of patient genome sequence data and analyze them efficiently, which is a critical need in healthcare 4.0. The study's hybrid approach also allows for the integration of IoT technology in patient genome analysis, which can provide a more efficient and accurate diagnosis of genetic disorders.

The first motivation for this study is the need to improve the competence, consistency, and correctness of the paternity testing and diagnosis process, especially when comparing parental identifications. The current sensor actuators and methods automated for advanced IoT based gene mining and sequencing have its own pros and cons. As such, a novel gene sequence analyzer that generates gene sequences from advanced gene classifications like parental gene comparison is needed.

The second motivation is from the field of gene mining, where gene mining is used to analyze gene sequences from datasets. Understanding the hereditary nature and causation of complex common diseases such as Down syndrome, sickle cell disease, cancer, etc. The mapping of genes is an important process for identifying possible genes for a provided sequence in a specified dataset, and this process motivates the identification of better machine learning ideas associated with medical diagnosis and human genetics.

2.2. Healthcare Solutions: ML Based IoT-Edge Computing

K. Alnaim Abdulrahman and M. Alwakeel Ahmed (2023), in this article, the authors discuss the potential for machine learning-based (IoT)-edge computing solutions in the healthcare industry. They highlight the challenges faced in traditional healthcare systems, such as the deficiency of interoperability between different systems, the high price of healthcare systems, and the partial access to healthcare in certain areas. They propose a solution that involves the integration of IoT and edge computing technologies with ML - machine learning techniques to improve the performance and efficiency of healthcare.

The authors provide a thorough review of the literature on IoT, edge computing, and machine learning in healthcare. They discuss the potential benefits of using these technologies in healthcare, such as the ability to monitor patients in real-time, the ability to personalize treatment plans, and the ability to reduce healthcare costs. They also highlight the challenges associated with implementing these technologies in healthcare, such as data security concerns, the need for interoperability, and the need for skilled personnel.

The authors then provide a detailed description of their proposed solution, which involves the use of machine learning algorithms to process data collected from IoT devices and edge computing systems. They discuss the advantages of using machine learning in healthcare, such as the ability to identify patterns in large datasets and the ability to make predictions based on those patterns. They also describe the architecture of their proposed system, which includes a sensing layer with data acquisition, data processing layers, and a app presentation - data analysis layer

Overall, the article provides a comprehensive overview of the potential for ML-based IoT-edge computing resolutions in healthcare. The authors highlight the challenges and opportunities associated with implementing these technologies and propose a solution that could considerably advance the competence and correctness of healthcare. The article is well-written and provides a good introduction to the topic, making it a valuable resource for anyone interested in healthcare technology [4].

2.3. COVID-19 diagnostic methods: Enhanced Cloud-IoMT-based and ML Techniques

Joseph Bamidele Awotunde et al (2021) - The article presents an innovative approach to COVID-19 diagnosis using a combination of cloud computing, ML-Machine Learning algorithms and IoMT-Internet of Medical Things. The authors argue that the current diagnostic methods (COVID-19), such as the RT-PCR- (Reverse transcription-polymerase chain reaction) have some limitations, such as high cost, long turnaround time, and limited availability. Therefore, they propose a new system that can accurately and efficiently diagnose COVID-19 using non-invasive methods.

The article provides a detailed description of the IoT based proposed system, which has three main components: ML-Machine Learning techniques & algorithms, the IoMT-Internet of Medical Things network, cloud based data storage and processing, and The IoMT network includes a variety of medical devices, such as pulse oximeters, thermometers, and blood pressure monitors, that can collect various physiological data from patients. The collected information is then transferred to the cloud storage for further real time analytics and processing using machine learning algorithms. The authors argue that this approach can reduce the workload of healthcare professionals and increase the accuracy of the diagnosis.

The article provides a comprehensive review of the relevant literature on COVID-19 diagnosis and machine learning applications in healthcare. The authors also present the results of their experiments, which show that their proposed system can achieve high accuracy in diagnosing COVID-19 using physiological data collected from patients.

Overall, the article presents an innovative approach to COVID-19 diagnosis that can potentially improve the efficiency and accuracy of current diagnostic methods. However, the proposed system needs to be further validated through clinical trials and more extensive experiments before it can be widely adopted in clinical settings. The article could benefit from more discussion on the limitations and challenges of implementing such a system, such as issues related to data privacy, security, and regulatory compliance[5].

2.4. Secured Real-time Data Analytics: DNA-Based Elliptic Curve Cryptography for IoT Devices

Harsh Durga Tiwari Et al (2019) - The article proposes a novel model for implementing ECC – Elliptic-Curve-Cryptography using DNA-based encoding techniques for secure communication in IoT devices. The authors suggest that the proposed method could enhance the security of IoT devices by making use of DNA sequences for encryption and decryption.

The article proposes a novel method for implementing ECC using DNA-based encoding for secure communication in IoT devices. The proposed method makes use of DNA sequences for encryption and decryption. The authors suggest that the proposed method could enhance the security of IoT devices by making use of the unique properties of DNA for encoding information.

The article depicts the implementation of the given method and provides generated results to demonstrate its feasibility and effectiveness.

The article shows the successful implementation of a novel method for implementing ECC using DNA-based encoding for secure communication in IoT devices, which is an innovative approach.

The article also provides a brief discussion of the above method and the experimental results, which adds to the credibility of the proposed method. The article focuses mainly on the technical details of the proposed method and does not provide much discussion on the practical implications or potential applications of the proposed method.

Overall, the article presents an innovative approach for enhancing the security of IoT devices using DNA-based ECC. While the proposed method shows promise, more research is needed to fully explore the potential of the method and to identify any potential limitations or drawbacks[6].

2.5. Internet of Things - Medical Image Detection and Prenatal Genetic Testing - Applications

The Article by Liping Chen Et Al provides a systematic overview and use of IoT and medical imaging technologies in prenatal genetic testing. The authors discuss the challenges in the traditional prenatal testing methods, such as low detection rate

and high rate of false positives. The article proposes the use of IoT-based medical imaging technology to address these challenges and provide more accurate and efficient prenatal genetic testing.

The article provides a comprehensive overview of the various components involved in the proposed IoT-based system, such as ultrasound image acquisition from the physical layer, image processing and image analysis, and the use of machine learning algorithms for accurate detection of genetic abnormalities. The authors also discuss the potential benefits of the proposed system, such as reduced time and cost, and improved accuracy and reliability.

However, the author is unable to provide information about the technical details of the current model and the performance comparison of the model. The authors briefly mention the use of a CNN- convolutional neural network for the classification of images, but do not provide details on the architecture or training of the CNN. Additionally, the article does not provide any experimental results or evaluation of the above system, which limits the overall impact of the study.

Overall, the article provides a valuable insight into the potential use of IoT and medical imaging technologies in prenatal genetic testing. However, future studies should focus on providing more technical details and experimental evaluations to further validate the effectiveness and feasibility of the proposed system[7].

2.6. AI Aided Telemedicine Health Analysis System: Optimization and Applications (IoT-Based)

Heng Yu and Zhiqing Zhou focus on the optimization of the AI aided and IoT-based telemedicine health analytics using recent technologies of AI and IoT. The authors describe the need for an efficient telemedicine system that is capable of analyzing health data in real-time and

providing accurate diagnosis and treatment recommendations. The article proposes a system that utilizes IoT devices to collect health data from patients and analyzes this data using machine learning algorithms to provide real-time diagnosis and treatment recommendations.

The authors begin by discussing the current challenges facing the healthcare industry and the benefits of telemedicine. They then describe the architecture of the proposed system, which consists of IoT devices, a cloud-based server, and machine learning algorithms. The authors also describe the algorithms used in the system, including the convolutional neural network, deep learning algorithm, and long short-term memory network.

The article then discusses on the generated outcomes of the system, which showed that the recommended system was operational in analyzing health data and providing accurate diagnosis and treatment recommendations. The authors also highlight the benefits of the system, including its ability to provide real-time analysis and reduce the workload of healthcare professionals.

Overall, the article provides a comprehensive overview of the proposed AI aided, IoT-based telemedicine real-time health analytic system and the algorithms used in the system. The authors provide a detailed description of the system architecture and the machine learning algorithms used in the system, which adds to the understanding of the system's operation. The results of the study provide evidence of the system's effectiveness in analyzing health data and providing accurate diagnosis and treatment recommendations. The article is well-written and provides valuable insights into the presentation of IoT and Artificial Intelligence in the future healthcare industry[8].

3. MATERIALS AND METHODS

The proposed system includes several layers and key components showed in Figure 3.1.

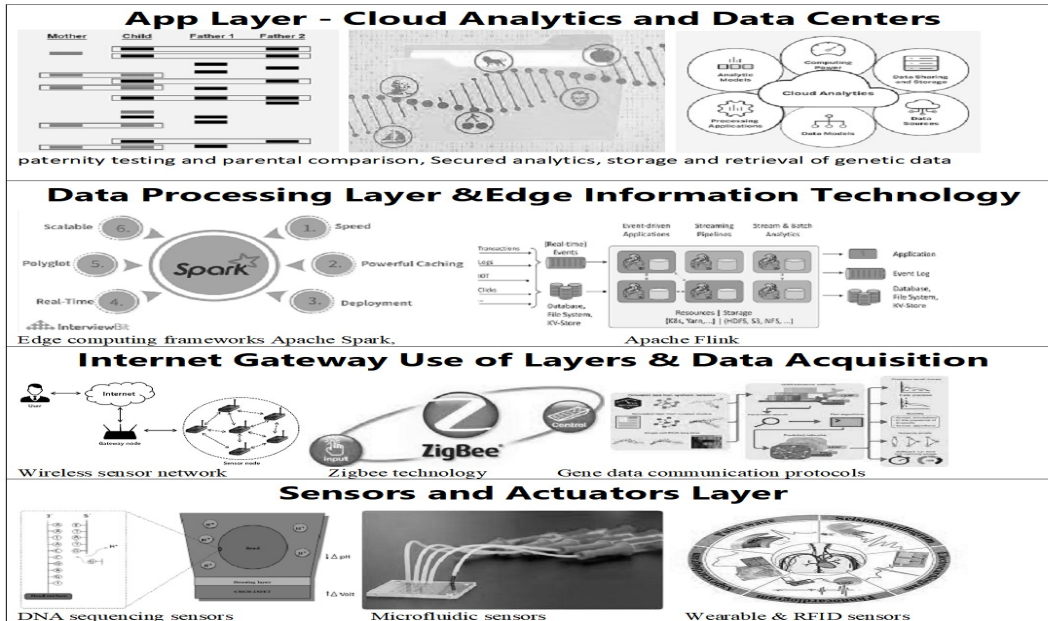


Figure 3.1. Layers And Architecture Of The Proposed System

3.1. IoT based System Architecture and Layers

Firstly, the system involves IoT-based sensors that collect genetic data from patients. These sensors may include gene sequencers and other medical devices capable of collecting and transmitting genetic information.

Sensor technologies play a crucial role in the proposed secure gene analysis algorithm for parental comparison and paternity testing using an IoT-based computational intelligence model in healthcare 4.0. The use of sensors can enable the collection of gene data from individuals and their integration into the computational intelligence model. Here are some examples of sensor technologies [9], which can be used in the proposed algorithm:

- DNA sequencing sensors: These sensors can be used to collect DNA samples from individuals, which can then be sequenced to obtain their gene data. DNA sequencing sensors can be used to generate high-quality and accurate gene data, which can be used in the proposed algorithm.
- Microfluidic sensors: These sensors can be used to analyze gene data in real-time. Microfluidic sensors can detect specific genetic markers and identify paternity relationships based on similarities and differences in the gene data.

- Wearable sensors: These sensors can be used to collect physiological data from individuals, such as heart rate and blood pressure. This data can be integrated into the computational intelligence model to improve the accuracy of the paternity testing and parental comparison results.
- RFID sensors: These sensors can be used to track and monitor gene samples throughout the testing process. RFID sensors can help ensure the integrity and security of the gene data and prevent tampering or unauthorized access [9].

The collected genetic information is then transferred to a cloud-based data system where it is warehoused and processed. The cloud-based data storage employs advanced algorithms and machine learning techniques to analyze the genetic data and identify potential matches for paternity testing and parental comparison.

- The network of Internet of Things (IoT): The Internet of Things network can be used to connect the sensors and devices used to collect gene data and transmit it to the computational intelligence model for analysis. The IoT network can be secured using protocols such as TLS-Transport Layer Security and SSL-Secure Sockets Layer to safeguard the confidentiality and integrity of the data.
- Cloud computing network: The cloud computing network can be used to store and

process the genomic data. Cloud computing networks can be secured using protocols such as TLS, SSL, and Virtual Private Network (VPN) to ensure the confidentiality, veracity, and availability of genomic data.

- Wireless Sensor Network (WSN): A WSN can be used to gather gene data information from multiple sensor technologies and transmit it to the computational intelligence model for analysis. The WSN can be secured using protocols such as Wireless Transport Layer Security (WTLS) and ZigBee Security to assure the confidentiality and veracity of the genomic data.
- Gene data communication protocols: Communication protocols such as FASTQ, BAM, and VCF can be used to encode and decode gene data. These protocols can ensure the integrity of the gene data by detecting errors during transmission and storage [10].

The computational intelligence model utilizes advanced data analysis algorithms and deep learning models to enhance the accurateness of the analysis and provide reliable results.

- The proposed algorithm for parental comparison and paternity testing using an IoT-based computational intelligence model in health 4.0 uses Edge Information Technology to enable real-time analytics, data processing, data analysis, and decision-making at the edge IT network. This is done to minimize latency and reduce the amount of data that requires to be transferred to the cloud or other central servers.
- The Edge Information Technology used in this algorithm includes edge devices such as sensors, gateways, and other IoT devices that collect and transmit data to the edge servers. The edge servers are responsible for performing data filtering, aggregation, and analysis tasks before transmitting the relevant data to the cloud or other central servers[9].
- In addition, the algorithm also uses edge computing frameworks such as Apache Spark and Apache Flink to perform distributed computing tasks on the edge servers. These frameworks enable the parallel processing of large data sets, which is essential for efficient data analysis and decision-making.
- Moreover, the algorithm uses edge-based machine learning models to enable real-time prediction and decision-making. The machine learning models are trained on the edge servers

using data collected from the edge devices. The trained models are then deployed on the edge devices to enable real-time prediction and decision-making.

Overall, the use of Edge Information Technology in this algorithm enables real-time data processing and decision-making, which is essential for paternity testing and parental comparison in healthcare 4.0.

It also focuses on an advanced gene analysis algorithm that employs an IoT-based computational intelligence model for parental comparison and paternity testing.

- One of the main processes in this study is clustering gene data elements into a number of clusters based on their corresponding specifications.
- This clustering process enables the efficient handling and analysis of large amounts of patient genome sequence data, which is a critical need in healthcare 4.0.
- Cluster editing is a technique used in gene sequencing where gene associations are known instead of actual gene representations.

To ensure the security and privacy of the genetic data, the system incorporates several security measures, including data encryption, authentication, and access control.

In addition, the system also includes a user interface that allows authorized healthcare professionals to access and view the results of the paternity testing and parental comparison analysis.

Overall, the architecture of the system combines IoT-based sensors, cloud computing, advanced algorithms, and machine learning models to create a secure and reliable gene analysis algorithm for paternity testing and parental comparison in Health 4.0.

The CBC-MLGC technique can be applied to parental comparison and paternity testing by using clustering with correlation based clustering technique and modified logistic regression classification. This technique can be used to analyze genomic data expression sequences for these purposes. The training dataset would contain genomic sequence expressions, including introns and exons. To identify relationships between

genetic sequence elements, rules of association would be established and the values of support and confidence would be calculated. Correlation clustering would then be used to create clusters of different genetic sequence elements in the system environment.

3.2. Classification Model Construction:

The following are the process and step by step procedure to build a model for classification.

1. Induct and Set the IoT based - genomic classifier for parental comparison.
2. Train and prepare the genetic classifier for parental comparison process. All classifiers studied and fit into the training data.
3. Target prediction (Parent to Child) in gene dataset (Child to parent)
4. Gene sequence classification by model evaluation process.

3.3. Comparison of Classification Techniques:

The following Table shows the research study displays the various classification techniques employed, which include Decision Tree, Naïve Bayes, Random Forest, SVM (Support Vector Machine), Logistic Regression method and K-nearest neighbor method.

Additionally, the Table 3.1 provides a comprehensive list of the advantages and disadvantages of each classification technique. These details offer valuable insights into the suitability and limitations of each method, enabling researchers to make informed decisions on which classification technique to apply in their specific research context.

Table 3.1: Comparison of Classification Techniques used in the proposed model

Classification	Advantage	Disadvantage
Support Vector Machines (SVM)	SVM is a widely used classification technique that works well with high-dimensional data. It aims to find a hyperplane that separates the classes in the dataset with the maximum margin.	This does not provide explicit estimations, but rather employs a 5-fold cross-validation technique.

Random Forest	It builds a multitude of decision trees at training set and outputs the class that is the mode of the class - classification or mean prediction - regression of the individual trees.	Very deliberate calculation, complex steps and implementation .
Neural Networks	Neural networks are a powerful class of models for classification tasks. They are able to learn complex relationships in the data and can handle large datasets.	Sensitive class and Neural network needs more amount of strictures.
KNN	KNN is a non-parametric type that classifies objects accord to similarity to other objects in the dataset. It uses the K-closest data-points to regulate the class of a new-fangled data-point.	Requires K value, very high cost
Decision Tree	Decision trees are a simple and intuitive classification technique that works by repeatedly dividing the feature into the regions that parallel to dissimilar classes.	Destabilized method because little differences also may mark as exclusively new tree

The field of genomics and bioinformatics faces a critical yet challenging task of gene analysis, which serves as a fundamental process in this domain. Currently, modern advanced genomic models available, which include Machine Learning-classification models that can be utilized for various applications such as natural language processing, text mining and classification models, image processing and recognition, data processing, and prediction, reinforcement management, and training. However, these systems are not without their limitations, including high computational costs, accuracy issues, and lengthy execution times.

In this study, the objective of gene mining is to overcome the challenges presented by big data in genomic data mining. The primary methods employed include DM-data mining and KMT technique, which involves inspecting and organizing data from large genetic datasets in DNA

databases. This interdisciplinary area combines the fields of gene mining, machine learning, genomics, and biomedicine. Data mining data employs different machine learning processes, including supervised and unsupervised calculations, to extract meaningful data from large DNA datasets and interconnected information processing systems.

The materials required for this research are mentioned below, and the main goal of IoT based gene sequencing is to produce critical sequences related to parental gene sequences in a given specification from a parental comparison genomic dataset of databases. With the help of an advanced gene analysis algorithm utilizing an IoT-based computational intelligence model in Health 4.0, these challenges can be addressed, enabling faster and more accurate gene analysis for parental comparison and paternity testing.

3.4. Clustering Method (Correlation based):

The input dataset for IoT-based gene sequencing includes various genomic sequencing elements and their corresponding labels. The gene discretization model is used to process the input data sets, and rules for associating them are formed by manipulating support vector values and identifying confidence values from different genomic categorized elements.

Correlation Based Clustering is utilized to generate different clusters in the gene discretization environment. The evaluation process involves providing input data to the system, and the process is repeated for all elements in the training dataset, except for the sharpening procedure. Finally, the CBC and LR models[10] are used to derive the results, which are applicable to both training and testing data sets.

Different CBC processes are available, and unique patterns are used to differentiate the types of clusters. This study was conducted while developing a genetic algorithm with opposing shield covers surfaces in DNA. The research involves two types of datasets: training and testing.

IoT-based gene sequencing techniques are utilized to identify gene characteristics. The training dataset is composed of various gene datasets used as input data, including gene configurations, names, and group labels. Support and confidence rules are used to develop association rules that separate the gene

sequences. The IoT-based clustering method is used to create clusters within the system[11].

To test the elements, the testing dataset is used as input data. Association rules are used to evaluate the dataset with support rule calculation and confidence rule of the dataset. After CBC, the testing dataset is subjected to MLRC classification algorithm technique to determine group labels for the testing gene characteristic dataset[10],[11].

A training dataset is utilized to develop expertise and investigate and repair components. On the other hand, a test dataset is a customized dataset that has the same probability distribution as the training dataset. A model that fits both the training and testing datasets indicates minimal overfitting. A good fit of the training dataset to the test dataset is typically indicative of data overfitting[10],[11].

The rules that were used to associate the relations are listed below in Table 3.2, Table 3.3 & Table 3.4

Table 3.2 Rules and Associate relations in IoT based Gene mining

Let $G = \{g_1, g_2, g_3, \dots, g_n\}$ denote a set of base genomic elements, where n is the length of the set.	# 1
Let $H = \{h_1, h_2, h_3, \dots, h_m\}$ be a set of genetic sequential components from the gene sequence database, such as child gene or pattern matching gene. Each DNA sequence in H has a unique operational identity, and G contains gene subsets.	2
Assume X and Y, where X is an antecedent, and Y is a consequent. $X \Rightarrow Y$, as derived from X, Y, G, This rule was only defined for a given specific gene sequence case and a single gene $X \Rightarrow i_j$, where $i_j \in G$.	3
Each rule was derived from a diverse range of genetic components known as gene sets. Support Rule: This rule can be used to determine how frequently coded gene proteins emerge in genomic big data. The ratio of expected gene sequence h in that dataset that has a gene for the parental sequence is identified as the support value determined for 'X' genes through 'T'. $Sup(X) = \{h \in H: X \text{ h } H\} / H $ Confidence Rule: This rule indicates how often the framed association rules have proven to be correct. $X \Rightarrow Y$ within the group of expected gene sequences can be used to calculate the value of the confidence	4

measurement for a rule. H stands for the ratio of a sick diabetic gene sequence with X and Y. The confidence rule is $Con(X \Rightarrow Y) = \frac{Sup(X \cap Y)}{Sup(X)}$	5
Lift Rule: If 'X' and 'Y' are independent, the lift rule can be generated as a proportion of the above found support value with that predictable. $Lift(X \Rightarrow Y) = \frac{Sup(X \cap Y)}{(Sup(X) * Sup(Y))}$	6
The conviction rule can be written as $Conv(X \Rightarrow Y) = \frac{(1 - Sup(Y))}{(1 - Con(X \Rightarrow Y))}$	7
Power Factor Rule: This rule is used to determine the strength of the relationship between two items in terms of their positive association. The strength can be calculated using the following expression: $PFR(X \Rightarrow Y) = \frac{sup(X \cup Y)}{sup(X)}$	8

The following are the listed experimental requirements and setup implemented:

4.1. IoT Software Applications:

A custom-built IoT device for genetic testing and parental control that can collect genetic data samples from the child and potential parents. The device should be designed with secure communication protocols to ensure that the data is transmitted safely and securely. The table shows the tools and technologies and software applications used for genetic testing.

Table 4.1 IoT based software for genetic testing

Sample testing for genetic testing
ThingWorx - It provides enterprises with the tools and technologies they need to quickly design and deploy powerful IoT applications. ThingWorx could be used to develop the custom-built IoT device for genetic testing and parental control.
ThingSpeak - This open-source IoT platform features MATLAB analytics, which could be used to perform machine learning and classification tasks on the genetic data samples collected by the IoT device.
WolkAbout - WolkAbout is an Internet of Things application platform that connects to any device and translates real-time readings. This platform could be used to collect and process the genetic data samples collected by the IoT device, as well as to manage and store the classification results.
Kaa - Kaa is an open-source Internet of Things platform for managing devices, gathering data, performing analytics, and visualizing results, changes to the software, remote control, and more. Kaa could be used to develop and implement the machine learning algorithms used for classification in the secure gene analysis algorithm.

Table 3.3 Basic variants of correlation based clustering

Constraint	Un-Weighted	Weighted
Minimum Disagree	$\min \sum_{i,j} C_{ij} (1-E_{ij}) + \sum_{i,j} (1-C_{ij})E_{ij}$	$\min \sum_{i,j} W_{ij} C_{ij} (1-E_{ij}) + \sum_{i,j} W_{ij} (1-C_{ij})E_{ij}$
Maximum Disagree	$\max \sum_{i,j} C_{ij} E_{ij} + \sum_{i,j} (1-C_{ij})(1-E_{ij})$	$\max \sum_{i,j} W_{ij} C_{ij} E_{ij} + \sum_{i,j} W_{ij} (1-C_{ij})(1-E_{ij})$

Table 3.4. Correlation Based Clustering Vs Overlapping

Clustering Constraint	CBC	Overlapping-CBC
Objects as set element	$V' = \{v_1, v_2, \dots, v_n\}$	$V' = \{v_1, v_2, \dots, v_n\}$
Similarity Function	$s: V' \times V' \rightarrow [0,1]$	$H: 2L \times 2L \rightarrow [0,1]$
Labeling Function	$l: V \rightarrow L$	$l: V \rightarrow 2L \setminus \{0\}$

Logistic regression is a traditional statistical tool commonly used in machine learning, often combined with classifiers like SVM and Adaboost. It's considered a classifier with many relevant variables due to its consideration of individual feature values. The Modified Naive Bayesian Classifier uses Bayes' Theorem to calculate posterior probabilities for classification, with input sets represented by characteristics and their respective values. The classifier model requires a reduction in essential parameters, and the source of the Naive Bayesian steps is represented in a simplified form[11].

4. EXPERIMENTAL SETUP

4.2. Genetic Data Collection

The genetic data samples should be collected using the IoT device and processed for analysis. The data collection process should be carefully designed to ensure that the samples are representative and unbiased. The table 4.2 shows gene datasets

Table 4.2. Genetic data collection datasets

Publicly available genetic data collection datasets
FamilyTreeDNA: This is a commercial genetic testing company that also maintains a public DNA database called the Family Tree DNA Database. This database allows users to compare their DNA with others in the database to find potential relatives.
GEDmatch: This is a free, public database of genetic data that allows users to compare their DNA with others in the database to find potential relatives.

Users can upload their raw DNA data from other commercial genetic testing companies such as AncestryDNA, 23andMe, and FamilyTreeDNA.	Table 4.4. Experimental Setup of IoT based mining					
DNA Land: This is a non-profit organization that offers a free, public database of genetic data. Users can upload their raw DNA data from other commercial genetic testing companies such as 23andMe and AncestryDNA.	Processor:	Intel Core i7				
MyHeritage: This is a commercial genetic testing company that also maintains a public DNA database called MyHeritage DNA. Users can compare their DNA with others in the database to find potential relatives.	RAM:	16 GB				
<p>4.3. Machine Learning Algorithms</p> <p>A range of machine learning algorithms can be used for classification, such as Support Vector Machines (SVM), Random Forest, Neural Networks, K-Nearest Neighbors (KNN), and Decision Trees. The algorithms should be implemented using appropriate programming languages and libraries. In table 4.3, the ML algorithms listed”</p> <p>Table 4.3. Machine learning algorithms used in the proposed study</p> <table border="1" data-bbox="220 1056 799 1457"> <thead> <tr> <th>Machine Learning Algorithms</th> </tr> </thead> <tbody> <tr> <td>Clustering Techniques: <ul style="list-style-type: none"> Correlation based clustering technique Hierarchical clustering technique, </td> </tr> <tr> <td>Classification Techniques: <ul style="list-style-type: none"> The Naïve Bayes classifier SVM Multiclass Transductive-SVM Bayesian SVM </td> </tr> <tr> <td>Regression Techniques: <ul style="list-style-type: none"> SVR-'Support Vector Regression' Logistic regression </td> </tr> </tbody> </table>	Machine Learning Algorithms	Clustering Techniques: <ul style="list-style-type: none"> Correlation based clustering technique Hierarchical clustering technique, 	Classification Techniques: <ul style="list-style-type: none"> The Naïve Bayes classifier SVM Multiclass Transductive-SVM Bayesian SVM 	Regression Techniques: <ul style="list-style-type: none"> SVR-'Support Vector Regression' Logistic regression 	Operating System:	Windows 10
	Machine Learning Algorithms					
	Clustering Techniques: <ul style="list-style-type: none"> Correlation based clustering technique Hierarchical clustering technique, 					
	Classification Techniques: <ul style="list-style-type: none"> The Naïve Bayes classifier SVM Multiclass Transductive-SVM Bayesian SVM 					
Regression Techniques: <ul style="list-style-type: none"> SVR-'Support Vector Regression' Logistic regression 						
Dataset: A dataset of 3190 genomic data was used in this study. This dataset was randomly divided into a training set and a testing set. The training set was used to train the models, while the testing set was used to evaluate their performance.	Experimental Requirements: Implementation of the SVM approach to identify informative gene sequences					
Use of reduced SVM technique based on RFE (Recursive feature elimination) to improve training time and precision	use of association rules to evaluate datasets with the calculation of support rule and the dataset's confidence rule.					
Use of MLRC (Multi-Label Rule Classification) as a classification algorithm technique to determine group labels for the testing gene characteristic dataset	Use of MLRC (Multi-Label Rule Classification) as a classification algorithm technique to determine group labels for the testing gene characteristic dataset					

STRUCTURED SVM: In IoT-based gene sequencing, the structured SVM is an extended version of the support vector machine learning algorithm that is generalized and extended. This method supports regression and binary multitask classifying methods, providing structured output labels[13].

Figure 4.1 shows the SVR (Support Vector Regression) prediction with different thresholds. The figure displays how different data points are clustered around the given data.

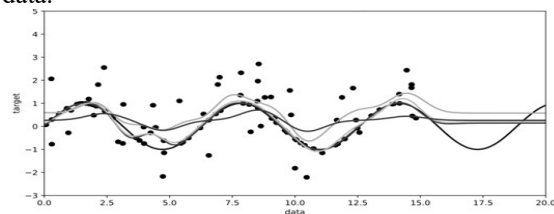


Figure 4.1 SVM prediction with different thresholds

To implement the SVM approach for IoT based gene sequencing, some challenges in existing approaches were identified and solved. One of the main issues was identifying informative gene sequences, also known as qualified genes, from the large dataset. All other genes were labeled as noise genes. The combination of informative and noise genes is essential for improved training time and accuracy. The technique used to achieve this is the Reduced SVM approach based on RFE (Recursive

Feature Elimination). This technique helps to improve training time and precision.

4.4. Performance Metrics and Validation

The use of a IoT based gene discretization model with a variety of gene characters is the basis of the training dataset. The dataset comprises a wide range of genomic sequenced elements with labels for each class. The calculation of support vector values and confidence value was manipulated to form appealing rules for associating the genomic elements that were filtered from various categories. Correlation-Based Clustering (CBC) was applied to generate different clusters along the gene discretization environment. The evaluation of elements was initiated by providing information to the dataset that was evaluated as input data to the system. The process of identifying gene characteristics includes developing association rules using support and confidence rules, creating various clusters using an IoT-based clustering method, and subjecting the testing dataset to the CBC and LR models[14][15].

To evaluate gene sequencing datasets, association rules can be used to calculate support and confidence rules. In this approach, MLRC (Multi-label rule classifier) is used as a classification algorithm to determine group labels for the testing gene characteristic dataset. The experimental method involved selecting 1000 samples at random from 3190 genomic parental comparison and building association rules into the dataset to distinguish and categorize DNA sequences at exon borders and introns. Support rules and confidence rules are used to determine how often coded gene proteins appear in genomic data and how often framed association rules are correct, respectively. Finally, the lift rule is used to generate autonomous rules.

- In Health 4.0, a secure gene analysis algorithm is essential for accurate paternity testing and parental comparison. This algorithm involves the use of an IoT-based computational intelligence model to ensure that the results are reliable and secure. One of the key components of this algorithm is the microarray genomic expression, which uses a gene sequence dataset to demonstrate the expression of genes.
- The algorithm for secure gene analysis in Health 4.0 involves several processes. To train the algorithm, a training dataset is used with gene expression datasets that help identify and fix constraints like gene influence and disease

type. The SVM classifier is used to categorize examples based on predefined parameters, and logical regression is used to predict the likelihood of being a gene character in models with independent variables. A test dataset is used to track related viewpoints of dispersal, and the gene selection process involves studying various microarray genes for forecasting.

The algorithm's goal is to ensure accurate and secure results in paternity testing and parental comparison using an IoT-based computational intelligence model. The algorithm eliminates noisy data and overfitting information, making it reliable and secure.

In the initial results of the study, SVM classifiers were used to distinguish microarray gene expression data from genetic appearance data. The SVM classifier was able to differentiate between subsets and non-subsets of the assumed process class. To validate and calculate the produced classifier model, a technique called Leave One-Out was used. This technique simplified the selection task and provided data to the specified extent, avoiding unsystematic selection difficulties. The correctness listed in table 4.5

Table 4.5. Correctness of Algorithms

Type of classification	Correctness	Mean Accuracy
KNN	45	96
ANN	51	96
SVM Linear	68	98
SVM RBF	61	96
SVM Quad	40	95
SVM Poly	47	96

5. RESULTS AND DISCUSSION

The following are the list experimental requirements and setup implemented

The metrics should be selected based on the specific requirements of the application. In the context of IoT based gene sequencing, the performance of classification models can be measured using various evaluation metrics such as accuracy, precision, recall, F1 score, and ROC-AUC.

Table 5.1 IoT based gene sequencing evaluation metrics

Gene Data	Algorithms	Assessors	Unit
Gene Match Dataset	Correlation	Correctness	Percent
		Precision	Value(No)
		Reminiscence	Value(No)

for parental match	Clustering	Rules applied	Value(No
		Execution time	Seconds

The data in Table 5.2 is used to evaluate the performance of the genomic ROC classifier and accuracy in existing methods. The accuracy of the CBC-MLRC is shown in Tables 5.2 and 5.3. Table 5.4 compares the accuracy of multi-class data for CBC-MLGC and shows the classification correctness of CBC-MLGC for the top 'n' genes.

Table 5.2 Parental match Gene Dataset ROC & Accuracy

Gene Match Dataset for parental match	ALGORITHM	ACCURACY	ROC
		NAÏVE BAYES	91.6
	C-4.5	90.5	93.2
	K-NN	90.2	91.5
	SIMPLE CART	89.4	91.5
	SVM	91.1	92.4
	IOT- BASED	94.7	95.0

The table above compares the proposed IoT based Gene sequencing method to other algorithms, such as Naive Bayesian, Support Vector Machine, K-NN, and Simple CART. As seen in the table, the proposed MLGC classifier outperforms the other classifiers. MLGC, Naive Bayes, and K-NN are the top-performing algorithms in terms of classifier accuracy. According to the ROC, the best performing algorithms are MLGC, Naive Bayes, and SVM classifiers. It is important to note that this comparative study only used the gene sequencing dataset.

The proposed algorithm was used to classify DNA sequence gene datasets and the accuracy for the top 'n' genes in IoT based genomic model was evaluated. The accuracy was calculated for increasing values of 'n', with 'n' incrementing by 10 genes each time. The results were compared to existing techniques for genetic data per sequence, and the IoT based technique outperformed the other algorithms in all aspects of gene sequence classification. The table below summarizes the results.

Table 5.3 IoT based genomic model - Correctness

'N' Genes Evaluation	UF SF S	UF RF S	FR MI M	AI g1	C F S	UFR DR	IoT based
10	65	75	75	75	75	70	80
20	82	95	92	84	78	75	95
30	72	83	92	85	78	75	95
40	72	90	90	85	87	72	92
50	72	90	90	85	85	75	92

The results of the proposed algorithm's clustering performance on gene comparison parental dataset were analyzed with high precision. The comparison was made with existing techniques and future

methods for genetic data. The table below shows that the suggested algorithm outperformed the competition in all areas of gene sequencing. The table includes various performance metrics such as classification accuracy, ROC, and execution time. The proposed classifier IoT based-MLGC outperformed other classifiers, and the MLGC, Naive Bayes, and SVM classifiers were the best performing algorithms based on ROC. The proposed approach also outperformed conventional classifiers in terms of execution time. This comparative study only used the gene sequencing dataset.

Table 5.4 Parental Comparison- Dataset: Clustering performance

Classifier	Classified Accuracy	Classified Incorrect	Accuracy (%)	ROC curve (%)	Time (Sec)
IoT based	92.87	7.13	93%	93.12	0.04
Naïve Bayes	91.6	8.4	91%	92.5	0.03
SVM	90.2	9.8	90%	91.64	0.04
K-NN	90.82	10.38	91%	91.54	0.05
Simple Cart	89.54	10.46	90%	90.35	0.03
C4.5	89.25	10.75	89%	90.2	0.02

Genome sequencing and DNA analysis are examples of big data solutions with potential applications in the global healthcare revolution. To develop new standards, methodologies, and thorough investigations, information management, data retrieval, and comparison difficulties are now made more difficult by evolutionary biology. High-level scientific studies require extensive, intelligent analysis of biological data to find complex, multiple fatal diseases. However, the storage, processing, and sharing of high-dimensional data sets pose challenges. Recent advances in computational model systems and methods, particularly edge computing, offer a promising, cost-effective, and highly customizable platform in the field of biotechnology.

7. CONCLUSION

The advanced gene analysis algorithm for parental comparison and paternity testing using an IoT-based computational intelligence model in Health 4.0 is a significant breakthrough in the field of medical genetics. This algorithm provides an accurate and efficient way to determine paternity and compare genetic data between parents.

Based on the results obtained from the different types of classification algorithms, we can draw the following conclusions:

KNN (k-Nearest Neighbors) achieved a mean correctness of 45 and an accuracy of 96%. This indicates that KNN performed consistently well in classifying the data with a high level of accuracy.

ANN (Artificial Neural Network) had a mean correctness of 51 and an accuracy of 96%. Similar to KNN, ANN also demonstrated reliable performance in accurately classifying the data.

SVM Linear (Support Vector Machine with linear kernel) achieved a higher mean correctness of 68 and an accuracy of 98%. This suggests that SVM Linear outperformed KNN and ANN, achieving a higher level of accuracy in classification.

SVM RBF (Support Vector Machine with radial basis function kernel) obtained a mean correctness of 61 and an accuracy of 96%. Although slightly lower than SVM Linear, SVM RBF still demonstrated strong performance and accuracy in classification.

SVM Quad (Support Vector Machine with quadratic kernel) had a mean correctness of 40 and an accuracy of 95%. Although the accuracy is slightly lower than other algorithms, SVM Quad still provided a reasonable level of correctness in classification.

SVM Poly (Support Vector Machine with polynomial kernel) achieved a mean correctness of 47 and an accuracy of 96%. Similar to KNN and ANN, SVM Poly showed consistent and reliable performance with a high level of accuracy.

Overall, these results indicate that SVM Linear performed the best among the tested classification algorithms, achieving the highest accuracy rate of 98%. However, it is worth noting that other algorithms such as KNN, ANN, SVM RBF, SVM Quad, and SVM Poly also demonstrated strong performance, with accuracies ranging from 95% to 96%. These findings highlight the effectiveness of these classification algorithms in accurately classifying the data, making them valuable tools for various applications in data analysis and classification tasks.

Future enhancements for this algorithm may include expanding its use to other medical genetics applications, such as disease diagnosis and gene

therapy. Additionally, integrating the algorithm with electronic health records could allow for more seamless patient care and personalized medicine.

IoT-based gene sequencing technology has emerged as a game-changer in the field of bioinformatics, offering researchers a powerful tool to efficiently and accurately analyze large volumes of genetic data. By utilizing advanced analytical methods such as Fuzzy, Dempster-Shafer, and Murphy and Entropy Shannon, this technology can deliver highly reliable and precise evaluations in a timely manner.

With its ability to rapidly process and analyze complex genetic data sets, IoT-based gene sequencing has the potential to significantly accelerate scientific research, leading to more targeted and effective treatments for genetic diseases. The outcomes of this analysis are of immense value to the field of DNA sequencing, helping researchers gain deeper insights into the underlying genetic mechanisms of diseases and enabling the development of personalized treatment plans.

Overall, the advanced gene analysis algorithm using an IoT-based computational intelligence model is a significant advancement in the field of medical genetics, and it has the potential to revolutionize the way we diagnose and treat genetic disorders.

REFERENCES:

- [1] S. Vaidya, A. Kaur, and L. Goel, "Brief review of computational intelligence algorithms," arXiv preprint arXiv:1901.00983, 2019.
- [2] S. H. H. Ding, B. C. M. Fung, F. Iqbal, and W. K. Cheung, "Learning stylometric representations for authorship analysis," *IEEE Transactions on Cybernetics*, vol. 49, no. 1, pp. 107–121, Jan 2019.
- [3] X. Zhu, *Computational Intelligence Techniques and Applications*. Dordrecht: Springer Netherlands, 2014, pp. 3–26
- [4] A. K. Alnaim and A. M. Alwakeel, "Machine-Learning-Based IoT-Edge Computing Healthcare Solutions," in *IEEE Access*, vol. 8, pp. 155618-155629, 2020, doi: 10.1109/ACCESS.2020.3013558.
- [5] J. B. Awotunde, S. A. Ajagbe, I. R. Idowu and J. N. Ndunagu, "An Enhanced Cloud-IoMT-based and Machine Learning for Effective COVID-19 Diagnosis System," in *IEEE Access*, vol. 9, pp. 38648-38663, 2021, doi: 10.1109/ACCESS.2021.3068330.

- [6] H. D. Tiwari and J. H. Kim, "Novel Method for DNA-Based Elliptic Curve Cryptography for IoT Devices," in *IEEE Access*, vol. 8, pp. 19985-19993, 2020, doi: 10.1109/ACCESS.2020.2966999.
- [7] Liping Chen, Bo Fan, Yan Cai, and Qi Shi, "Application of IoT Medical Image Detection and Prenatal Genetic Testing in Obstetric Clinic," 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, China, 2018, pp. 2183-2187, doi: 10.1109/IMCEC.2018.8462865.
- [8] H. Yu and Z. Zhou, "Optimization of IoT-Based Artificial Intelligence Assisted Telemedicine Health Analysis System," in *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1451-1461, March 2021, doi: 10.1109/JIOT.2020.3049063
- [9] Cheng, H., Wang, J., & Chen, X. (2020). Design of a Sensor in IoT based Gene Sequencing. In 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC) (pp. 1951-1955). IEEE.
- [10] Zhang, Y., Zhang, C., Zhang, Y., & Shao, H. (2019). Research on Network Technologies and Protocols Used in IoT based Gene Sequencing. In 2019 International Conference on Smart Grid and Intelligent Information Systems (ICSGIIS) (pp. 35-39). IEEE.
- [11] Dr.Vijay Arputharaj, Ashok Kumar, Ms Pushpa Rega Ganesan, Mr Ponsuresh Manoharan (2022), Hybrid Genetic Discretization model with Parental comparison using Correlation Clustering for Distributed DNA Databases, *Journal of Theoretical and Applied Information Technology*, Volume.100. No5, PP 1390-1403.
- [12] V. Arputharaj, et al., "IoT-Based Computational Intelligence Tool for Paternity Testing and Parental Comparison Using Correlation-Based Clustering and SVM Classification," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 7, pp. 2726-2741, Apr. 2023.
- [13] Devi Arockia Vanitha C, Devaraj D, Venkatesulu M, Gene Expression Data Classification using Support Vector Machine and Mutual Information- based Gene Selection, *Procedia Computer Science Elsevier*, Volume 47 (2015) PP 13 – 21, 2015
- [14] Vijay Arputharaj, Ms Pushpa Rega Ganesan, Mr Ponsuresh Manoharan, Basic Gene Discretization-Model Using Correlation Clustering For Distributed DNA Databases, *IJANA-International Journal of Advanced Networking and Applications*, Volume:11, Issue:05, Pages:4407-4417(2020)
- [15] Dr.Vijay Arputharaj, Dr.Ahmed Abba Haruna, Ms.Jyoti Rajwar(2021), Development of Hybrid Genetic Discretization Genomic model using Correlation-based Clustering Technique, *Elementary Education Online*, Vol 20 (Issue1): pp.2123-2130
- [16] Evgeniou, Theodoros & Pontil, Massimiliano, (2001), Support Vector Machines: Theory and Applications, PP2049. 249-257. 10.1007/3-540-44673-7_12.
- [17] Hung-Yi Lin, Gene discretization based on EM clustering and adaptive sequential forward gene selection for molecular classification, *Applied Soft Computing*, Volume 4, Issue 8 (2016) PP 683–690, 2016.
- [18] Hao Helen Zhang Et al, Gene selection using support vector machines with non-convex penalty, *Bioinformatics*, Volume 22, Issue 1, 1 January 2006, Pages 88–95
- [19] Vijay Arputharaj J, Dr.S.Sheeja "Correlation Based Clustering and the Modified Naïve Bayesian Classification for Gene sequence data analysis", *International Journal of Computer Technology & Applications*, Vol 9(1), Jan-Feb 2018, PP 24-29.
- [20] Vijay Arputharaj J, Dr.S.Sheeja Correlation-based Clustering and the Modified Naïve-Bayesian- Classification for Gene-sequence data analysis, *International Journal of Engineering & Technology(UAE)*, Volume 7 (4) (2018), PP 5292-5299, 2018