

Artificial Intelligence Approaches in Healthcare Informatics Toward Advanced Computation and Analysis



E.B. Priyanka^{1*}, S. Thangavel¹, R Mohanasundaram² and Shamala Subramaniam³

¹Department of Mechatronics Engineering, Kongu Engineering College, Perundurai, India

²School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamil Nadu, India

³Department of Communication Technology and Networks, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, Serdang 43400, Malaysia

Abstract:

Introduction: Automated Machine Learning or AutoML is a set of approaches and processes to make machine learning accessible for non-experts. AutoML can exhibit optimized enhancement of an existing model or suggest the best models for precise datasets. In the field of computerized Artificial Intelligence (AI), medical experts better utilize AI models with available encrypted information science ability.

Methods: This paper aims to characterize and summarize the stage-wise design of Automated Machine Learning (AutoML) analysis e-healthcare platform starting from the sensing layer and transmission to the cloud using IoT (Internet of Things). To support the AutoML concept, the Auto Weka2.0 package, which serves as the open-source software platform, holds the predominant priority for experimental analysis to generate statistical reports.

Results: To validate the entire framework, a case study on Glaucoma diagnosis using the AutoML concept is carried out, and its identification of best-fit model configuration rates is also presented. The Auto-ML built-in model possesses a higher influence factor to generate population-level statistics from the available individual patient histories.

Conclusion: Further, AutoML is integrated with the Closed-loop Healthcare Feature Store (CHFS) to support data analysts with an automated end-to-end ML pipeline to help clinical experts provide better medical examination through automated mode.

Keywords: Closed loop healthcare feature store, Data analytics, Medical datasets, AutoML, Glaucoma, AI models.

© 2024 The Author(s). Published by Bentham Open.

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: <https://creativecommons.org/licenses/by/4.0/legalcode>. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

*Address correspondence to this author at the Department of Mechatronics Engineering, Kongu Engineering College, Perundurai, India; E-mail: priyankabhaskaran1993@gmail.com

Cite as: Priyanka E, Thangavel S, Mohanasundaram R, Subramaniam S. Artificial Intelligence Approaches in Healthcare Informatics Toward Advanced Computation and Analysis. Open Biomed Eng J, 2024; 18: e18741207281491. <http://dx.doi.org/10.2174/0118741207281491240118060019>



Received: September 10, 2023

Revised: December 29, 2023

Accepted: January 09, 2024

Published: January 29, 2024



Send Orders for Reprints to reprints@benthamscience.net

1. INTRODUCTION

S-Health (Self Assessed-Health) is another type of medical care that is a subfield of e-health utilizing Electronic Health Records (EHR) and different factors from the framework of perceptive city to improve medical care [1]. The possibility of an improved well-being framework is that it utilizes all information coming from sensors on the body of patients and robots to help settle

on better choices and improve medical care by giving crisis reactions and paging specialists, medical attendants, and experts, as shown in Fig. (1). It can likewise uphold self-finding, checking, early recognition, and treatments [2]. Since the health care field inhibits various categories of examination health data, all the experts cannot process the database with minimum intervention. Hence, the pro-

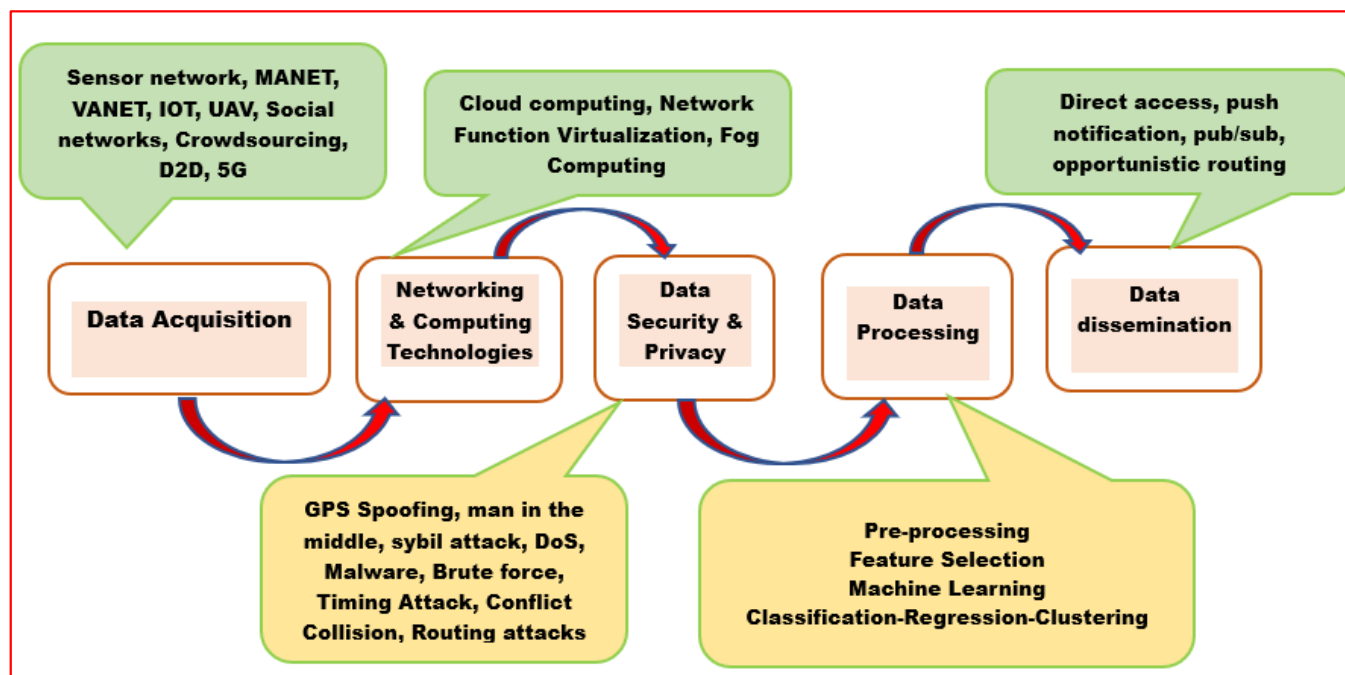


Fig. (1). Possible intercorrelation in the healthcare world and its threats.

posed Auto ML with Auto Weka package will better help the clinical experts to pre-process the data to get useful information using various ML approaches. The better-predicted model will be taken as the best-fit model followed by suggesting the cause rates and remedial precautions to be executed by the clinical experts to reduce the risk rates. Even Non-experts will be accessible to this Auto-ML platform to identify the cause and impact nature of the health status of the patients by providing the monitored datasets. The proposed platform with Machine Learning can support the medical informatics in the following aspects:

- Despite the obvious technological potential, machine learning technology has not yet gained the acceptance that it deserves in the field of medical diagnostics. It is difficult to anticipate, though, that this gap between technological potential and real-world application would persist for very long.

- Diagnostic tools based on machine learning will be utilized by doctors in the same way as any other tool at their disposal: as an additional source of potentially helpful data that enhances diagnostic precision. The doctor will nevertheless, as always, have the last say over whether to accept or reject this information.

- The amount of money individuals spend on different complementary medicine treatments is another indicator of the growing importance of complementary medicine in recent years. As more doctors become aware of the effectiveness and advantages of complementary medicine, they will require verification processes before they accept the benefits and provide permission to employ

complementary therapies. Because data analysis is transparent, machine learning can be especially useful in this process.

The improved well-being framework can be portrayed as a pipeline that incorporates information procurement, systems administration and figuring innovations, information security and protection, information handling, and information spread. Participatory administration, astute administration of assets, maintainability, and effective portability ensure the protection and security of the citizens [3, 4].

There is a crossing point between s-Health and e-Health (e-Health); e-Health can be characterized as “arising portable correspondences and organization innovations for medical services systems” [5, 6]. AI (ML) is a field that outgrew Artificial Intelligence (AI).

1.1. Comprehensive Literature Review

To secure user data, several state-of-the-art ML-based models have already been created to identify intrusion in network profiling data. Additionally, traditional ML models are not practical for solving an issue linked to large dimensionality data since they are effective at classifying tiny and low dimensional data [7, 8]. Additionally, because of problems with data imbalance, current approaches are unable to provide high performance, yet accuracy only accurately represents classes with high dispersion [9]. Additionally available as open source are ExploreKit and AutoLearn. Despite slight variations in how each method handles the feature transformations used for analysis and the feature selection phase, no method consistently performs better than the others. Using evolutionary

algorithmic techniques like genetic programming, the feature transformation can be emphasized more in e-health data mining. In their genetic programming solution for a feature engineering issue, Rahmani *et al.* [10] recommended the implementation of tree-based representation on the implicit feature selection as well as feature creation at the analytical formulation. Although this method yielded favorable results in several investigations, it also presents an unsteady resolution. On comparing with the expand-reduce strategy, their method did offer a marginal speedup. The feature engineering problem may also be approached using a hierarchical arrangement of transformations and reinforcement learning. It employs a greedy search method to gradually maximize the accuracy of the model as it explores different feature-building options hierarchically. To do this, all legitimate input characteristics are transformed using a directed acyclic transformation graph [11]. As a result, one can employ data-level alterations as logical building blocks for tracking performance over time, simulating a human trial-and-error approach. It also has the advantage of enabling transformation compositions. Later, the same authors presented a related approach that makes use of reinforcement learning [12], in which a reinforcement learning agent explores a transformation graph.

According to the findings of this study, information can be a tremendous asset for the healthcare sector, and the advantages of using medical data and analytics in various healthcare settings have raised awareness of healthcare informatics to a significant degree. The interdisciplinary approach to healthcare analytics poses some difficult problems and encourages a lot of creative solutions. So far, there has not been clear, well-organized research on the development of healthcare analytics. It is crucial to

undertake a full analysis of the healthcare analytics spectrum, from evolution concepts to the future direction of research, to gain a clear understanding of state-of-the-art technology. To close this gap, we use AutoML to highlight the many possibilities and difficulties faced by healthcare analytics and informatics in this study.

Furthermore, artificial intelligence and data analytics of the medical domain originated in three other important dimensions: the initial will be the health-medicine domain, followed by the second domain as online, and the final domain as data dimensions. To enhance the operation efficiency, quality, and cost, the utilization of Intelligent agents by taking information science contribution in terms of machine learning and neural networking satisfies the demand and expands the propagation of health services. Fig. (2) shows the contribution rate of supervised and unsupervised machine learning techniques in the e-healthcare world.

Further, Table 1 implies the merits and demerits of the impacts of Machine Learning Algorithms on healthcare informatics to take the digital world to a higher hierarchical strategy rate. Fig. (3) details the major Machine Learning Techniques (MLT) involved in healthcare data analysis to make an efficient decision-making system. The TPOT framework explicitly utilizes administered characterization administrators (for example, Irregular Forest, KNN, Logistic Regression, and so on), highlights preprocessing administrators (for example, Scalars, PCA, Polynomial Featurization, and so on), and includes choice administrators (for example Fluctuation Thresholders, RFE, and so on). To consolidate these administrators into an undeniable pipeline, they are treated as hereditary programming natives, and hereditary programming trees are developed [13].

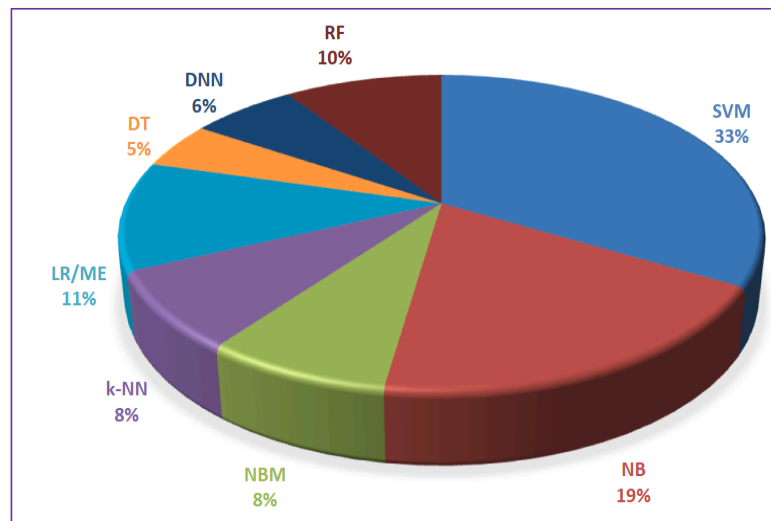


Fig. (2). Distribution of machine learning algorithms in healthcare informatics [47].

Abbreviations: RF-Random Forest, SVM-Support Vector Machine, NB-Nave Bayes, NBM-Naive Bayes Multinomial, K-NN-K-Nearest Neighbour, LR-Linear Regression, ME- Maximized Expectation, DT-Decision Tree, DNN-Deep Neural Network.

Table 1. Impacts of different machine learning algorithms in disease diagnosis.

Machine Learning Techniques	Disease Diagnosis Types	Merits	Demerits
Support Vector Machine (SVM) [32]	Heart Disease	Good performance timing and accuracy	Overfitting is high and robust to noise with pair-wise classification
Nave Bayes [33]	Diabetes	Better feature selections holding dissimilar attributes with minimum computation time	Requires large datasets to attain precise results and process entire streams, making model tedious
FT (Functional Tree) Algorithm [39]	Liver	Easy interpretation and fast prediction	Complex computation and uncertainty prediction are not feasible
Rough Set Theory (RS) [40]	Dengue	Data Significance is high and identifies hidden patterns with a minimal set of data	Imitation is high and unsuitable for complex datasets
Feedforward Neural Network [46]	Hepatitis	Adaptive learning with redundant information coding	Great Computation effort is needed to sort out overfitting; the model becomes a black box
CART (Classification and Regression Trees) [48]	Spinal disorder	Good computation with a minimum error rate	Requires continuous feedback to complete layer configuration

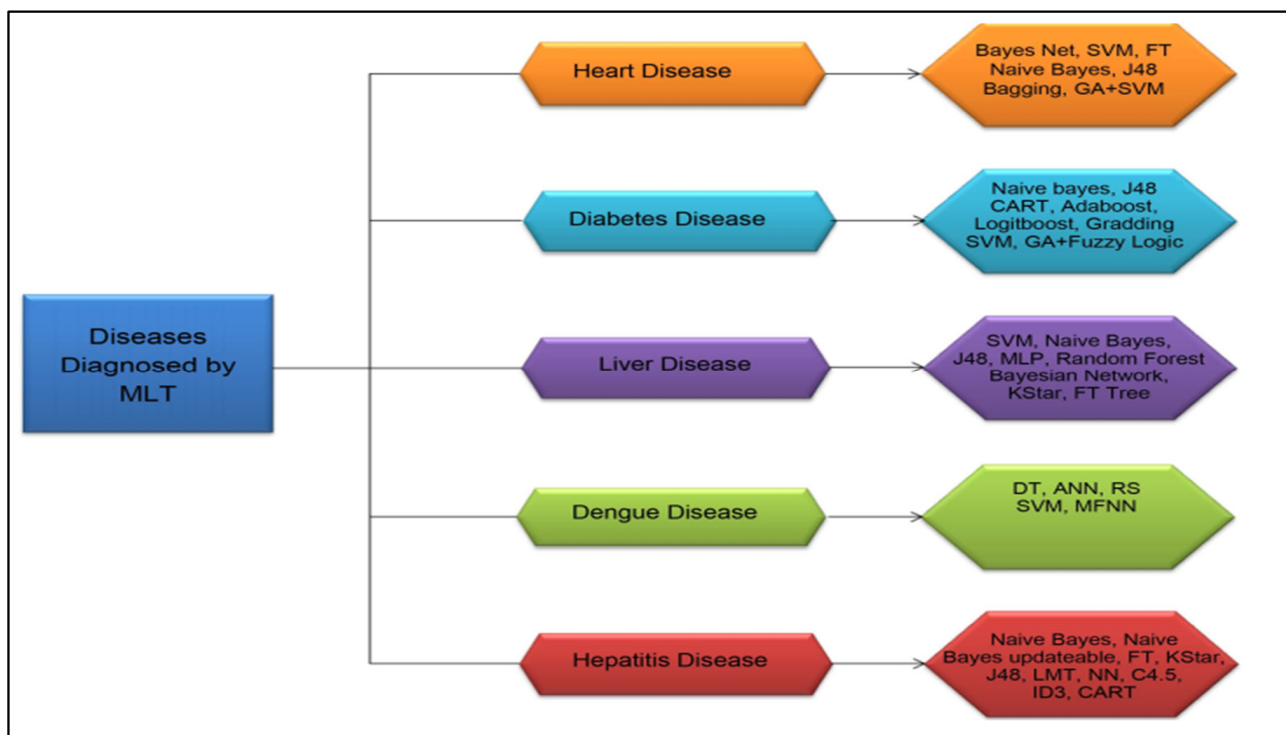


Fig. (3). Various Machine learning techniques are involved in healthcare informatics analysis streams [44]. [MLP-Multilayer Perceptron, ANN-Artificial Neural Network, MFNN-Mean Forward Neural Network, LMT-Logistic Model Tree].

Software agents are one of the most intriguing research paradigms for creating software applications in the healthcare industry, and they need to be able to sense both the real and virtual environment utilizing a variety of sensing devices. Agent-based computing systems in the healthcare industry have also been referred to as “the next significant revolution in healthcare service development through software.” Healthcare professionals may access and update the medical records of a patient utilizing portable devices in smart systems from any location. Additionally, integrated systems that leverage healthcare analytics are used to coordinate the many tasks necessary to deliver intelligent, interactive healthcare to the public. These intelligent systems combine several AI approaches

for a particular goal, among other goals falling under the e-health umbrella [14].

The proposed work concentrates on Automated ML and its characterization and possibilities in merging healthcare work and its computing platform. For real-time analysis and to improve the diagnostic application, automated machine learning (AutoML) searches for automatic selection and composition. It optimizes the machine learning models to attain robust performance on any type of dataset sample. The structure of the proposed paper is as follows: Section 1 outlines the healthcare informatics along with the contribution of Machine learning impacts. Section 2 describes the ML approaches

and existing architecture followed by a stage-wise IoT framework with an encrypted e-healthcare cloud with secured data transmission. It also outlines the advanced software tool package of WEKA 2.0 to support AutoML processing. Section 3 explains the feature engineering importance in the extraction and optimization of datasets to make predefined diagnostics to enhance the e-Health data warehouse. Section 4 enumerates the contribution of data mining using AutoML on the Glaucoma diagnostic with its classified outputs, followed by a summarization of results in the conclusion in Section 5.

2. METHODS

2.1. Existing Limitations of the E-healthcare System and its Frame Architecture

While a lack of high-quality data might be a barrier to the implementation of an AutoML system, the lack of transparency in these “black-box” AutoML systems is a considerably greater problem. Machine learning professionals will analyze the data through automated computation to make minor decisions within a short time. Users will be hesitant to utilize the outputs of AutoML in important applications if they lack confidence in the AutoML system, where interpretability and transparency of data analytics make the workflow a more complex and robust platform using advanced algorithms.

The fact that the present approaches to pipeline optimization using machine learning are ineffective for the kind of huge datasets that seem predominant in the biomedical setting is another explanation for the absence of AutoML in healthcare informatics. Pioneer researchers have presented dual software systems called PredicT-ML, which holds a tendency to automate machine learning for clinical big data to address major bio-mechanics issues. The implementation of AutoML in the healthcare industry is occasionally constrained by the temporal aggregation of clinical variables, which is a frequent property of clinical data in EHRs and is especially addressed by these systems [15]. Although these systems have intriguing designs, to our knowledge, they have not yet been used in clinical settings. The AutoPrognosis system is capable of handling many clinical data types, such as time-to-event and longitudinal data, and can also do feature preprocessing and imputation of missing data. Empirical research carried out by the creators of the system proved that the incorporation of the AutoML technique was better than the present clinical ratings. It has recently been shown that the AutoPrognosis technique improves the accuracy of risk prediction by correlating with the scoring systems by undertaking traditional emergency indicators. These two examples demonstrate how AutoML has the potential to be an essential tool for helping academics and medical practitioners with insightful knowledge to improve patient care [16].

Building a mathematical model based on measurable data is a regular occurrence in healthcare settings, including examination and patient health monitoring and control, with or without deeper experimentation [17]. Due

to the complex nature of huge data sample analysis, the machine learning issue in this field is frequently characterized by the availability of diverged high-dimensional training samples [18]. To convert the high-dimensional dataset into a lower-dimensional space, dimensionality reduction methods are typically needed [19, 20]. The recent experimentation unveils that AutoWEKA and ML-Plan are quite competitive compared to Auto-sklearn and ML-Plan [21]. The authors are now optimizing across pipelines concurrently utilizing techniques from the sci-kit-learn and WEKA packages. Auto stacker optimizes hyperparameters across hierarchically stacked machine learning models via an evolutionary algorithmic technique. It was influenced by the stacking method of group learning [22]. As a result, it is comparable to the TPOT strategy but takes advantage of ensemble learning. Autostacker is substantially quicker and performs on par with TPOT even though it does not preprocess data or choose or engineer features [23, 24]. Due to the inclusion of all the activities, the resultant pipeline is intrinsically very easy to understand.

The authors combine Monte Carlo tree searches (MCTS) and deep neural networks in their sequence modeling technique to address the optimization problem [25]. AlphaD3M performs magnitude order analysis on the measuring signal faster in calculations by incorporating the AutoML systems and self-forecasting by using the OpenML repository [26]. According to the empirical findings, this approach often outperforms Autosklearn. Deep learning-based picture inpainting techniques have recently demonstrated glaring benefits over currently used conventional techniques. The former can provide more aesthetically appealing picture structure and texture data [27, 28]. However, the most popular convolutional neural network techniques now in use frequently have issues with excessive color variation as well as texture loss and distortion in images [29]. The impact of picture inpainting is restricted to images with huge missing sections and low-resolution levels, and the majority of image inpainting algorithms frequently have issues with fuzzy images, texture distortion, and semantic inaccuracy [30].

2.1.1. Sensing Layer Framework

The sensing layer in the ML-associated healthcare applications will acquire and pre-process the data using Edge IT technologies for communication to the storage layer. For healthcare applications, ML is used to make proactive decision platforms. It provides a clear image of the many IoT device types that might be utilized to identify different diseases, such as Parkinson's, blood cancer, stroke, cardiac, and lung cancer. Several IoT sensors on the market can track the health of people. Biosensors play the most important function among them. IoT, which relies on biosensors, paves the way for the human race to transition to completely digitalized e-healthcare services [31]. IoT devices are utilized in several medical applications, including the detection of neurological activity, the prediction of heart disease and stroke, the

monitoring of glucose levels, the care of the elderly, the identification of lung cancer, clinical diagnostics, fall detection, and smart home care.

2.1.2. Location and Processing of Biosensors

The positioning of the biosensing plays a vital role in design architecture with processing technique and supporting interface for pre-emptive risk analysis. Modern smartphones are equipped with a multitude of built-in sensors, including magnetometers, barometers, accelerometers, gyroscopes, ambient light sensors, microphones, and more. Only a small percentage of the ever-expanding array of sensors included in smartphones have shown promise in applications connected to healthcare. These bio-sensors are designed with firmware processing and electronics functionalities on a micro or nano scale to process the data of patients for next-level examination by pushing to the cloud through smart gateways with a high-frequency switching spectrum. Few examples are

- Using a camera to record images to analyze the optical assay readout using resolution variations with CMOS technology
- Using a microphone, pulmonary functions are examined by incorporating MEMS technology with a range of 0.02-9 kHz and sensitivity from -30 to -41dB
- Using an accelerometer, step activity is measured with a sensitivity of 0.512 to 0.045 mg/LSB
- Using ECG, electrical potential through a digital form is acquired with a maximum resolution of DC/AC (microvolts) of 9-99 with six patches localized area on the chest.
- Using microfluidics to find the run assay by recapitulating the chemical assays through optical transparency of 69% optical transmittance rate of 399-699 nm

For the monitoring of cardiac and/or respiratory activity, as well as syndromes linked to their deficits, novel measurement methods and data analysis approaches have been developed. In particular, respiratory activity has been studied as a measure associated with sleep disorders. The scientists used a non-contact technique based on thermal imaging cameras to analyze respiration and body movements without disturbing people while they were sleeping. They conducted their analysis to evaluate the respiratory rate estimation performances. Additionally, a multi-sensor module to track breath activity by ECG and SpO₂ was suggested for patients with obstructive sleep apnea-hypopnea syndrome. For diagnosing obstructive sleep apnea, the hardware solution was integrated with cutting-edge neural network classification techniques [33]. Another non-contact approach for calculating the respiratory rate has been put forth, and it involves using an RGB camera built into a laptop and a bespoke algorithm to compare the results to a reference value.

Four healthy participants were used to test a unique wearable system based on fiber optic sensors for detecting pulse waves and measuring blood pressure. The data taken from a healthy person who holds normal blood pressure and glucose levels with a good cholesterol rate by maintaining proper Body Mass Index weight served as a predominant database for comparison to identify the deviations. The sensor integration highlighted how crucial it is to increase pulse wave signal accuracy and repeatability for their solution [34].

Information consolidated from various resources through Mobile Ad hoc Web of Things (IoT), 5G devices, and Device-to-Device (D2D) is grouped and formatted with packet transmission for efficient transport [35]. The systems administration and figuring advancements are applied to the information assembled in the information securing stage; since this information is a cycle complex, we need a worker or a figuring innovation where the information can be prepared, for example, distributed computing or haze figuring [36]. The information preparation includes preprocessing to eliminate the commotion and clean the information, highlight extraction to discover the significant highlights, and ML strategy to handle the information and play out the necessary assignment, for example, grouping/bunching. The direct access or pop-up messages with discerning steering on the huge dataset of the yield result are liable for the cloud utility rate with self-reconfigurability. Automized E-Health systems are designed for circumstances found in daily life, even if small wireless modules with ML devices cannot offer the same level of precision, resulting in rigorous medical decisions. It is predicted that this class of Auto-ML-enabled smart modules will be within the confines of the medical lab and enhance the vast market for smart consumer devices and digital services [37].

2.2. Development of Encrypted Health Record Data in E-cloud Using IOT

A significant quantity of healthcare data is now being dispersed over the network by utilizing the Internet of Things (IoT) and cloud healthcare technologies. In an IoT and cloud environment, it is crucial to guarantee the privacy and security of the diagnosis of patients. To safeguard their data from unauthorized access, IoT services need security protocols and guidelines. Enterprise software has undeniably expanded significantly across a wide range of industries and government agencies, creating and disseminating enormous amounts of data (including multimedia data). Choices and suggestions are made using these data. The cloud computing paradigm is a result of the fusion of many data sources and the ensuing massive accumulation of data. One of the key benefits of cloud computing is that it enables cross-sector data access, real-time calculation, and cross-jurisdictional decision-making using cloud resources. However, data flow between nodes, users, or

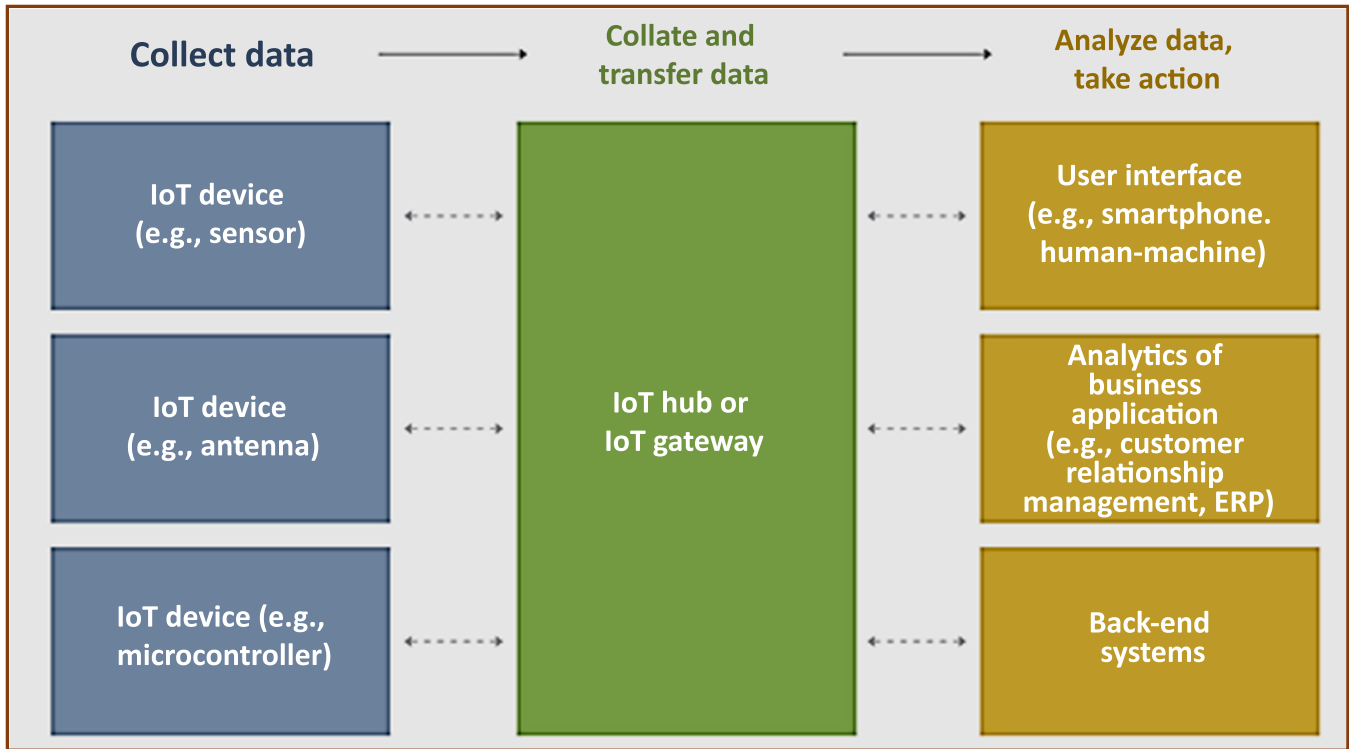


Fig. (4). IoT integration in healthcare data management.

across the cloud platform is extremely insecure unless a strong protective solution is offered.

To reduce test duplication, hospitals share medical data, including individual reports, clinical outcomes, etc. Patient treatment is expedited via the interchange of medical information. Thus, the modern era of e-health incorporated to promote s-health requirement is for the exchange of medical data with proper encryption techniques to ensure data safety and access. The suggested model carries out the cryptographic system. Fig. (4) shows the IoT-incorporated platform to access and retrieve e-health data from the healthcare private cloud.

As shown in Fig. (4), physicians, nurses, chemists, radiologists, lab technicians, and patients make up the healthcare ecosystem. The organization of the medical record is aided by cloud computing at several levels of the healthcare system. By linking several health information management systems with laboratories, pharmacies, radiology, and other components of the healthcare ecosystem, cloud computing is a promising and growing technology for those who use it. Data security and privacy concerns are the key roadblocks and significant challenges preventing the cloud computing industry from expanding quickly. Due to the cloud service profitability, health care requires software as a service with the public cloud utility. Though the users prefer the private cloud, the safety associated with the retrieval of e-health data still does not meet the limit of the encryption threshold rate. Hence, it is vital to perform feature engineering on the received

datasets to make proactive decisions in the diagnostic phase [38].

Due to the sheer size of big data, processing is split up and done on several nodes. Furthermore, Hadoop/MapReduce and other cloud-based open-source technologies have encouraged the application of big data analytics in the medical field. Although the mathematics and concepts of big data analytics tools and conventional analytics tools are similar, their user interfaces differ greatly. They are largely open-source development tools and platforms that have been developed ad hoc; hence, they lack the support and user-friendliness of vendor-driven proprietary products. This data must be gathered to do big data analytics. The second component involves processing or transforming the “raw” data, which has several possible outcomes. One solution is to use web services in combination with a service-oriented architecture approach.

Encryption and decryption procedures make up the cryptographic scheme. The plain text T is separated into odd components F_{odd} and even part F_{even} , during the encryption process. F_{odd} is encrypted with the AES (Advanced Encryption System) using a secret public key. F_{even} is encrypted using the RSA (Rivest-Shamir-Adleman) algorithm using the secret public key m. Eqs. (1 and 2 below can be used to quantitatively model the encryption process:

$$\vec{F}_t = T(Ke_{ff}F_{even} + Ke_{hk}\vec{h}_{t-1} + b_{\vec{k}}) \quad (1)$$

$$\vec{F}_t = T(\text{Ke}_{ft}F_{odd} + \text{Ke}_{hk}\vec{h}_{t+1} + b_k) \quad (2)$$

To boost security, the private key x used in the side decryption procedure of the receiver is encrypted using the AES technique before being transferred to the recipient as shown in Eq. (3).

$$Q_t = F_{even}e_{hf}\vec{h}_t + F_{odd}e_{hf}\vec{h}_t + b_k \quad (3)$$

As a result, it divides the total amount of supplied text data into F_{odd} and F_{even} . T_{odd} was encrypted using the AES cryptosystem, while F_{even} was encrypted using RSA. Detailed encryption support with IoT and cloud platforms is given in Fig. (5). Given the need for computational performance, 64-bit RSA was investigated; however, 256-bit AES was used in Eq. (4).

$$F(T, k) = -\sum_{i=1}^n f^{(i)} \log(F_{odd}^{(i)} + (1 - x^{(i)})\log(1 - F_{even}^{(i)})) \quad (4)$$

Low-bit-size RSA has frequently been criticized for being less resilient; however, this criticism only applies to RSA when used as a stand-alone encryption scheme. Its integration with AES-256 can assist in achieving two objectives. The combination of AES-RSA as a cryptographic technique can first prevent the possibility of simple attacks (which is achievable with independent

encryption). Second, taking low-bit size into account can prevent unnecessary processing which will eventually make it reliable to support real-time applications. The most used method of cryptography is data encryption. In this case, the data is utilized to be inserted or hidden within the image so that the attacker will not be able to get the data, while the authorized person will be given a key called the encryption/decryption key to obtain the data. As the key length rises linearly in this situation, it is difficult to discover a workable encryption/decryption key mostly because of the combination of values utilized to construct exponential key ranges.

The identity key length of the encryption/decryption key grows exponentially, making it impossible to use straightforward search techniques like linear or binary searching. According to the aforementioned paradigm, patient data will be encrypted before being stored in the cloud. The hospital systems access the encrypted patient symptom data of the cloud and compare it to the symptoms of already-existing patients, which are likewise in an encrypted form. The completely homomorphic encrypted enabled model selects the diagnostic and executes the operation on the encryption system and encrypted data. Applying the AutoML idea to the final outputs can help determine the optimal diagnosis.

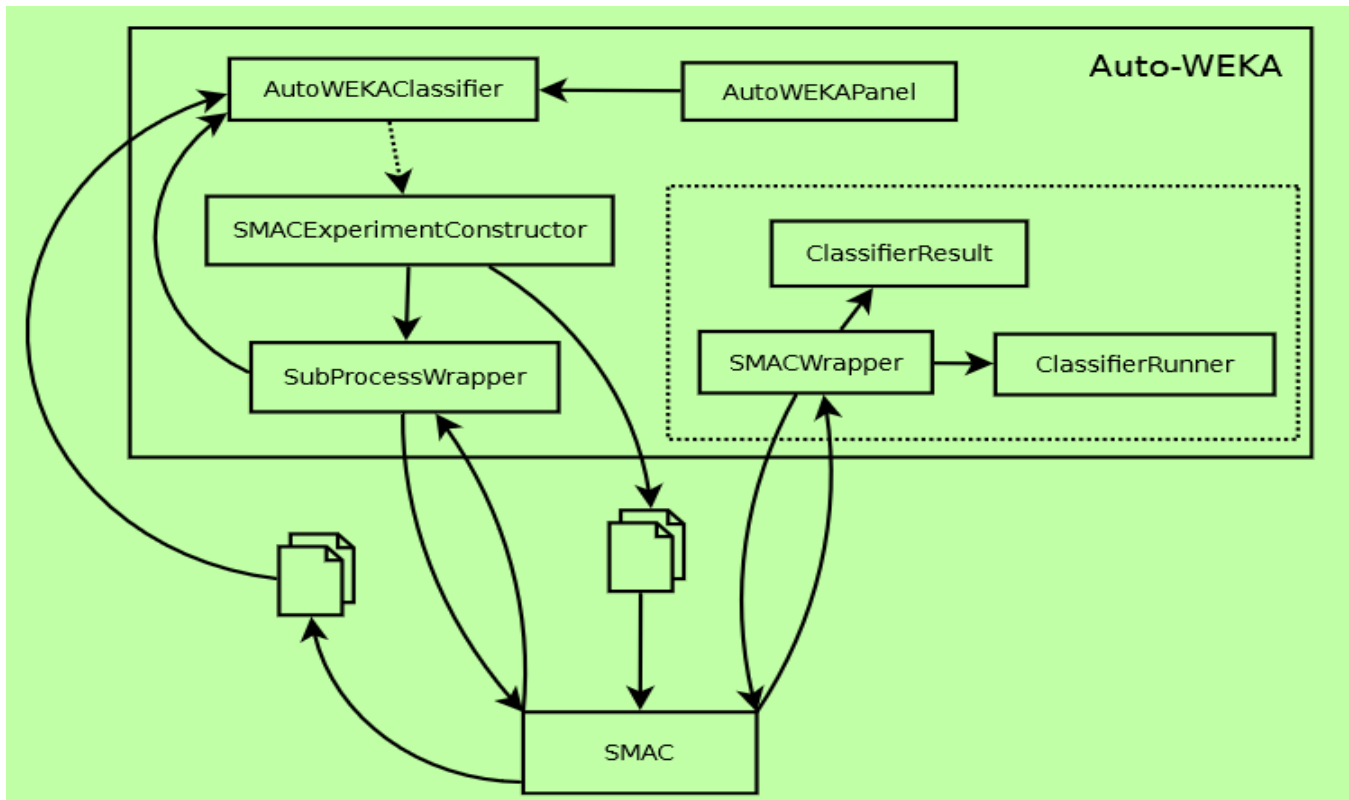


Fig. (5). High-level internal structure of auto weka2.0 architecture.

3. SCOPE AND CHARACTERIZATION OF AUTO-ML WITH AUTO WEKA 2.0 PACKAGES

A high-level overview of the internal structure of AutoWEKA is given in Fig. (4). The user interface of AutoWEKA is specified in the AutoWEKA classifier and has more than 10,000 lines of code. There are classes for AutoWEKA Panel, where the latter describes the components of the customized calls of the AutoWEKA Classifier from the AutoWEKA panel and new user choices and advertising additional justifications ought to be given there. The SMAC Experiment Constructor is used to invoke the SMAC optimization tool and classes for the sub-process wrapper. The first creates a collection of files that include the information required for SMAC to function, including the data, the parameter space definition, the training budget, and other details as outlined in Fig. (6).

These files are produced by the system-temporary files for temporary are kept in a folder called autoweke[random number]. The SMAC is called by Sub-Process Wrapper as an external process and directed to these files. During the optimization, SMAC then invokes the SMACWrapper class from the AutoWEKA evaluation procedure for the parameters and performance of a classifier. The SMACWrapper calibrates and assesses the classifier using the specified parameters using the classifier result created by the Classifier Runner class. This knowledge is returned to SMAC. Once SMAC has finished optimizing, control is returned to AutoWEKA. It uses the data directly transmitted from SMAC through its SMAC results files, which are output and parsed. This data is employed to choose the ideal classifier and its settings, which are then provided to a user.

There is an expanding appropriation of well-being data innovation, for example, electronic e-health records (EHRs), across clinical claims to fame and settings [40]. For instance, in the United States, 96% of medical clinics and a comparable level of outpatient settings have embraced EHRs for their regular clinical consideration. The measure of wellbeing-related information in EHR frameworks develops dramatically with the expanding appropriation of well-being data innovation. In any case, about 80% of all well-being information is caught as text (*e.g.*, family ancestry, progress outlines, radiology reports, and so on), which makes them less usable for medical services practice and research needs. The quickly developing volumes of text information require new methodologies for information handling and examination. Common language preparation (CLP) is a field that joins software engineering, computational etymology, and well-being skills to create computerized ways to deal with separate significance from clinical story information. Customarily, CLP was executed by utilizing either rule-based or AI draws near.

Rule-based CLP frameworks are frequently founded on pre-characterized vocabularies that incorporate complex clinical rationale. For instance, a CLP framework called "Clinical Text Extraction, Reasoning, and Mapping System" (CTERMS) was later used to recognize twisted

data in clinical notes. The injury ID calculation depended on an altered extensive clinical jargon of more than 200 terms portraying wounds and a set of decisions that were utilized to precisely recognize wound size and other wound attributes. Making an extensive jargon of wound-related terms was one of the most detailed pieces of the task [41]. The venture group utilized a blend of writing surveys, standard wordings, and a different scope of clinical notes to distinguish the up-and-comer words that were then explored by the investigation specialists.

The least difficult and most credulous hyperparameter enhancement methodologies occupy a predominant place in the big space of data analytics. Framework computation will be the least complex approach to accomplish meta-parameter streamlining, as it is a basic fundamental technique in which the client determines a limited arrangement of qualities for individual hyperparameters and, afterward, assesses the Cartesian validation of rough sets. This calculation unmistakably experiences the scourge of reductional, as the computation time develops dramatically tedious over the size of the setup space which impacts reasonable decisions given its enormous time necessities. A basic option, in contrast to network search, is irregular inquiry. Arbitrary pursuit depends on examining hyperparameter setups from a client-determined arrangement of hyperparameter values until a specific financial plan for the pursuit is depleted. Regardless of the way that irregular pursuit does not search the same number of setups as framework search, it has still appeared to perform observationally in a way that is better than network search [42]. While matrix search and arbitrary inquiry are basic methods that can fill in as a helpful standard, neither of these strategies utilizes past execution assessments and are, along these lines, wasteful at investigating the hunt space.

Another class of hyperparameter enhancement techniques is "advancement from tests" strategies, monitored for inquiries that relatively produce rough clusters dependent on the past execution of earlier arrangements. Two famous instances of these sorts of techniques are swarm streamlining (SS) and developmental calculations, the two of which are propelled by natural practices. SS is enlivened by how natural networks interface at the industry and societal level with technological upgradation. SS works by refreshing the mining space at every cycle by exploring the arrangement of the best singular setups and looking through the nearby computational designs in secondary emphases. Interestingly, developmental calculations are propelled by organic development and executed by keeping a populace (arrangement space) and progressing the populace by incorporating changes (little annoyances) and hybrid (joining singular answers for) to acquire an "age" of better arrangements. Probably the best usage of populace-oriented techniques is the Covariance Lattice Adaption Advancement Methodology (CLA-AM), which tests the multivariate Gaussian appropriation levels whose mean covariance is refreshed in each iteration.

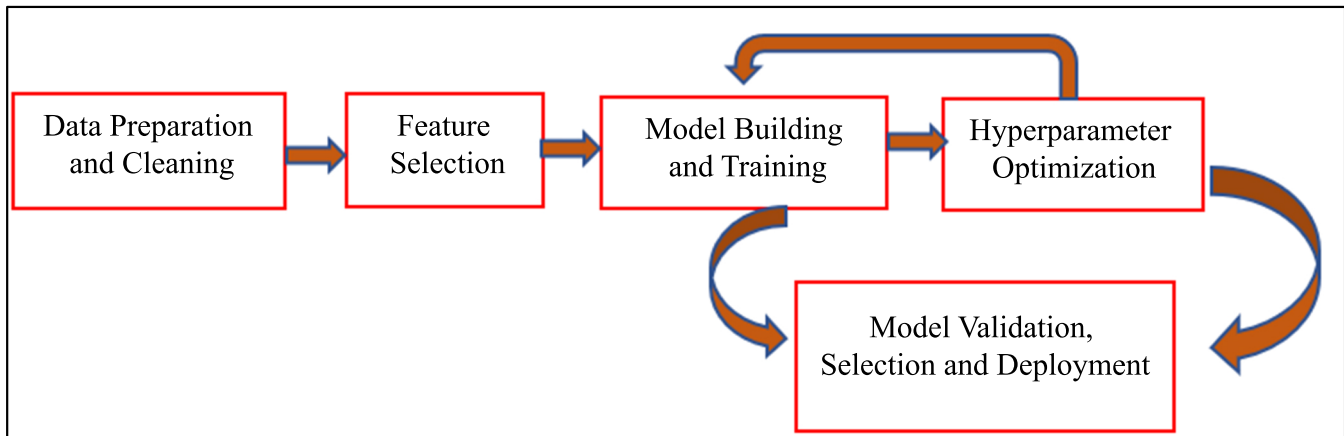


Fig. (6). Pipeline characterization of auto AI in healthcare mining [43].

Recently, Bayesian improvement has arisen as a novelty enhancement structure on the AutoAI framework. Bayesian enhancement follows the probabilistic analysis with iterative calculation by correlating two primary segments: a substitute evaluated model and a procurement work. Bayesian streamlining initiates the model with the probabilistic proxy model assembly, ordinarily as a Gaussian measure or resembling a tree-based model utilized to plan the distinctive analysis on the hyperparameter designs with the equal weightage of the vulnerability. Utilizing this substitute model, a securing capacity is then characterized to decide the expected utility of guaranteed setup and subsequently balance investigation and abuse during the inquiry cycle. Bayesian improvement has a solid hypothetical defense and has appeared to function efficiently, making it the most broadly utilized technique for streamlining hyperparameters [43]. There are a few diverse open-source executions of Bayesian streamlining. Fig. (7) describes the pipeline network associated with AutoAI in the healthcare world.

While a significant part of the exploration done on hyperparameter streamlining is noteworthy, the vast majority of these techniques are as yet restricted in proficiency inside the setting of huge biomedical information conditions. The measure of time needed to look for ideal hyperparameters develops quickly as the setup space, dimensionality, and information focuses develop. To conquer a portion of these impediments, an

execution of Bayesian streamlining is proposed that employs reformist testing to build an apparatus to empower medical services scientists to perform AI all alone. Likewise, the BOHB (Robust and Efficient Hyperparameter Optimization at Scale) calculation has joined the standards of Bayesian streamlining and Hyperband. This outlaw inquiry methodology utilizes progressive dividing, trying to assemble a stronger and more proficient hyperparameter analyzer that works at scale [44]. The updated toll package of Auto-WEKA, a program initiated to assist these manipulators by automating and systematically evaluating the combined space of the learning algorithms used by WEKA and their respective modern Bayesian optimization technique, hyperparameter choices are made to maximize performance. Due to its strong integration with WEKA, the present software platform overcomes the conventional issues of identifying the best machine learning for their specified datasets. As with any other learning algorithm, Auto-WEKA1 solves this issue by treating the entirety of WEKA as a single, high parameter with the use of Bayesian optimization and metric machine learning framework to identify a robust indicator for a certain dataset. It specifically takes into account the unified learning space of WEKA. Algorithm A is equal to A(1), A(k), and the hyperparameter spaces that go with the various certainty levels of $\Lambda(k)$ and (k) .

The activation function associated with Sequential model-based Bayesian optimization is given below:

$$EI_{y^*}(x) = \frac{\gamma y^* \ell(x) - \ell(x) \int_{-\infty}^{y^*} p(y) dy}{\gamma \ell(x) + (1-\gamma)g(x)} \propto \left(\gamma + \frac{g(x)}{\ell(x)}(1-\gamma) \right)^{-1} \quad (5)$$

Sequential model-based Bayesian optimization, such as black-box optimization problems, can be solved iteratively using optimization-based techniques. $\theta_i, f(\theta_i); n-1, i=1$ chooses the following n to be evaluated using this model (trading off the investigation of exploitation of known-to-be-good regions vs. exploration of unknown portions of

space) and assesses $f(n)$. In contrast, Gaussian process models-based Bayesian optimization is known to be effective for difficulties in low dimensions using numerical hyperparameters. It has been demonstrated that tree-based models perform better for high-dimensional, structured challenges that are partially discrete.

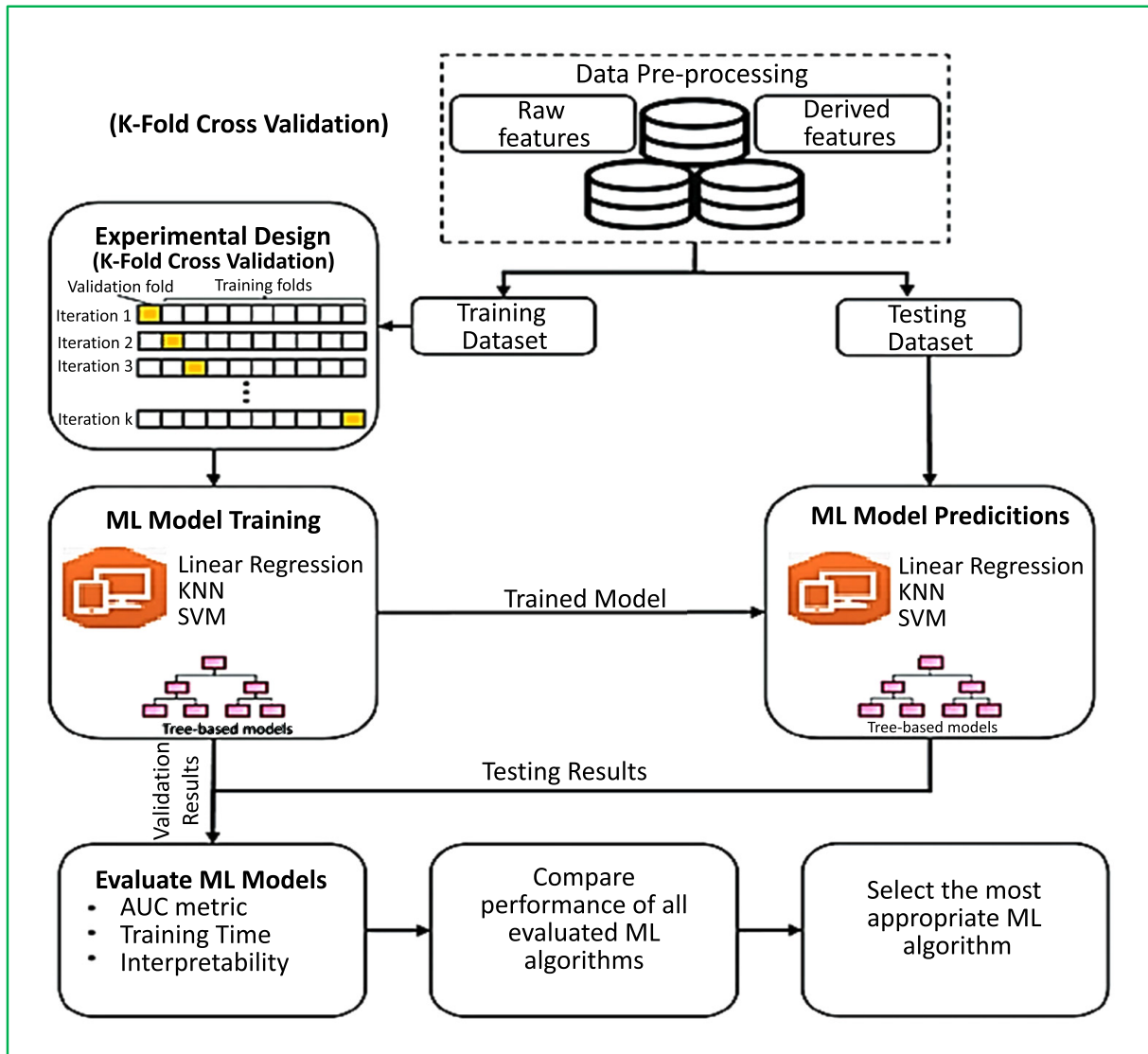


Fig. (7). Schematic view of the healthcare data analysis with AutoML.

There are two distinct distributions for the hyperparameters of Equation 1: one when the objective function value seems to be lesser than the threshold, $l(x)$, and another scenario of holding greater than the threshold, $p(x)$. On pointing, the Tree-structured Parzen Estimator operates. The objective function is then used to evaluate the ideal hyperparameters for the $p(z | x)$ surrogate function. There is no expectation that the hyperparameters x will result in any improvement if $p(z | x)$ nearly reaches zero anywhere when $z < z^*$.

4. PROPOSED AUTO-AI FRAMEWORK WITH IN-BUILT EXTRACTION AND OPTIMIZATION

Robotizing a partial task demanding human ability will permit the medical world to more quickly assemble, approve, and send AI arrangements, and consequently more promptly receive the rewards of improving the nature of medical services for patients. Fig. (8) shows the

detailed schematic representation of the processing of the e-health world by AutoML. The AutoML challenges center around comprehending administered machine learning issues with no human mediation, given some computational limitations. These computational limitations were somewhat distinctive across the difficulties. However, they typically incorporate a time interval (~18 min for preparing and validating) and memory use restrictions. The initialization of three rounds of holding (18GB RAM) is followed by 56GB towards the completion [45].

A portion of the rivalries incorporated a GPU track, yet entries on the track variations and updates were inadequate. The objective of ML background arrangement of black box concepts is to make a complete cycle that eliminates the greater part of the necessities for human skill in machine figuring to resolve the cluster issues space

information. There are difficulties in three-level hierarchy stages, each with marginally diverse issue details and datasets, and there is, as of now, a continuous test for worldly social information, which just resembles a future task for PC vision. Take into account an instance X with M attributes, each of which has a range-level axe. Eq. (6) creates the grouping of instance characteristics M into the primary optimum value of the cost function.

$$F_m = \sum_{x=1}^I \sum_{y=1}^W j_{xy}^m \|v_x - d_y\|^2 + M_{xy} \quad (6)$$

To fix its boundary constraints Eq. (7) is formulated to pointing mean difference between the point sets of Inner criteria and output criteria.

$$P_{xy} = \sum_{l \in W, i, x \neq j} \frac{1}{t_{xj} + 3} (1 - n_{yj})^r \|v_x - d_y\|^2 \quad (7)$$

The time function of the AutoML execution given in Eq. (8) is based on the productive best-fit model by taking base model and the computational model along with Euclidean estimation for the segregation of data.

$$t_{xy} = \frac{1}{\sum_{v=1}^u \left(\frac{\|v_x - d_y\|^2 + P_{xy}}{\|v_x - d_l\|^2 + P_{xl}} \right)} \quad (8)$$

The prior assessment optimized value validation of feature extracted function will be formed from the correlation of data in the source with threshold values, as represented in Eq. (9).

$$F'_{xy} = \sum_{x=1}^I \sum_{y=1}^W j_{xy}^m \sum_{x \neq j} n_{xj} (1 - v_{yj}) (1 - t(v_j, d_y)) \quad (9)$$

Eq. (10) will be a final channelized model function with a better-optimized extracted model to feed as input to the AutoML platform for better processing of testing data.

$$f_{xy} = \frac{1}{\sum_{l=1}^m \left(\frac{(1 - P(v_x, d_y)) + \sum_{x \neq j} j_{j \in N_y} t_{xj} (1 - n_{yj})^m (1 - P(v_j, d_y))}{(1 - P(v_x, d_l)) + \sum_{x \neq j} j_{j \in N_y} t_{xj} (1 - n_{yj})^m (1 - P(v_j, d_l))} \right)^{\frac{1}{z-1}}} \quad (10)$$

Notwithstanding a request for an AI framework that will be applicable "off-the-rack" by a master-slave configuration, the requirement for AutoML frameworks can be accompanied to handle a wide range of undertakings [46]. In this present scenario, consideration of existing frameworks of pipeline streamlining agents will be compared, and the best feature abstraction will be optimized. Every pipeline streamlining agent achieves at least a single assignment to help robotize the AI cycle. The principal pipeline analyzer undertaken for the present analysis includes WEKA, an AutoML framework dependent on the mainstream AI and information analysis stage. Auto-WEKA was the main AutoML framework that considered the issue of choosing an AI calculation and streamlining its hyperparameters; an issue which the makers named the joined calculation determination and meta parameter streamlining (MS) issue. The MS issue

can be seen as only a progressive meta-parameter streamlining issue, where the decision of logic analysis on a hyperparameter will be initiated. The MS issue can be the more officially characterized platform to execute algorithm upgradation to enhance the model processing on the rough sets.

The Auto-WEKA stage uses the SMAC (social, mobile, analytics, and cloud) streamlining calculation and tackles the CASH issue (Credit Accumulation Stream Hour) utilizing the students furthermore, include selectors actualized in the WEKA stage. This meta-learning analysis initiates by first assessing a bunch of meta-features for 220 diverse datasets in the pool of OpenML warehouse. At this point, Bayesian streamlining decides and starts to access the machine learning pipeline towards solid observational execution on each dataset [47]. In addition, when given another dataset, the calculation processes its meta-features, positions the wide range of various datasets by L1 distance estimation on the D variable in the meta-feature platform, and chooses the put-away AI algorithm by holding k = 32 closest datasets on the assessment before beginning Bayesian advancement. The subsequent enhancement attained *via* Auto-sklearn seems to be robotized outfit development models evaluated during enhancement. As opposed to disposing of these models, Autosklearn stores them and utilizes a post-preparing technique to build an outfit out of them.

5. RESULTS AND DISCUSSIONS

5.1. Case Study on Data Mining for Glaucoma

This section implicates the effects and impacts of glaucoma, which gets severe due the influence of diabetes all over the world. To resolve the data analysis on healthcare informatics, many ML techniques are examined, which is given in Table 2. Each algorithm shows different accuracy levels on the undertaken datasets of different extreme levels. Maternal risks of Diabetes Mellitus include polyhydramnios, pre-eclampsia, uterine atony, postpartum hemorrhage, infection, and progression of retinopathy and glaucoma, which are the main symptoms of morbidity. In Table 2, open-source availability and software of different datasets are utilized by comparing the several structured datasets undergone with supervised and unsupervised algorithms followed by neural network architecture.

Glaucoma, if left untreated, can result in vision loss and blindness and is twice as likely to develop in those with diabetes. Glaucoma risk can be raised by diabetic retinopathy, the most prevalent form of diabetic eye disease and a consequence of diabetes. Changes in blood glucose levels can deteriorate and harm the blood vessels of the retina in diabetic retinopathy. Glaucoma may eventually result from the early stage of Diabetes, affecting the retinal area gradually. The prevalent theory is that damaged retinal blood vessels can result in neurovascular glaucoma, which is characterized by aberrant blood vessel growth in the eye. Since due to diabetes, eyesight problems will be the initial stage of diabetes mellitus occurrence, this section details the

detection of retinal features causing glaucoma. The proposed work consists of three various data processing and validation stages. It includes the first step, Region of Interest (ROI) extraction, segmentation, followed by the Feature extraction stage finally, the classification stage. The block diagram of the validated methodology in the data analysis is shown in Fig. (9).

Three different segmentation techniques, such as the Gradient-based active contour method, Level set, and K-

means clustering-based segmentation method, are proposed for the extraction of optic disc and optic cup. The retinal pictures have been taken in an RGB mode by a fundus camera. A sample image taken for analysis is given in Fig. (10). The assessed area around the distinguished, most splendid point is to be chosen for the introductory optic circle district as ROI. After investigating the whole assortment of fundus pictures, a square of size 360x360 pixels with the most splendid pixel as the middle point is chosen to consider as ROI.

Table 2. Various machine learning implementations in diabetes diagnosis.

Machine Learning Techniques	Author	Year	Disease	Dataset Resources	Implied Software Platform	Obtained Accuracy
Hybrid Techniques (GA [Genetic Algorithm] + SVM)	Dhayanand <i>et al.</i> ,	2015	Liver	UCI	WEKA	89.05%
Naïve Bayes	Ephzibah EP <i>et al.</i>	2011	Diabetes	Pima Indian Diabetes dataset from UCI	R project Version 2.12.2	78.56%
(Support Vector Machine) SVM	Kumari VA <i>et al.</i>	2013	Diabetes	ILPD from UCI	MATLAB with SQL server	75.92%
DT (Decision Tree)	Sarwar <i>et al.</i>	2012	Diabetes Type-2	Pima Indian Diabetes dataset	MATLAB 2010a	69.99%
RF (Random Forest)	Iyer <i>et al.</i>	2015	Diabetes	Different Sectors of Society in India	R software	73.33%
LR (Linear Regression)	Tan <i>et al.</i>	2019	Hepatitis	Public Health Department of Selangor Stat	WEKA	77.84%
DNN (Deep Neural Network)	Sen <i>et al.</i>	2014	Diabetes	Diabetic Research Institute in Chennai	MATLAB neural network Toolbox	86.37%

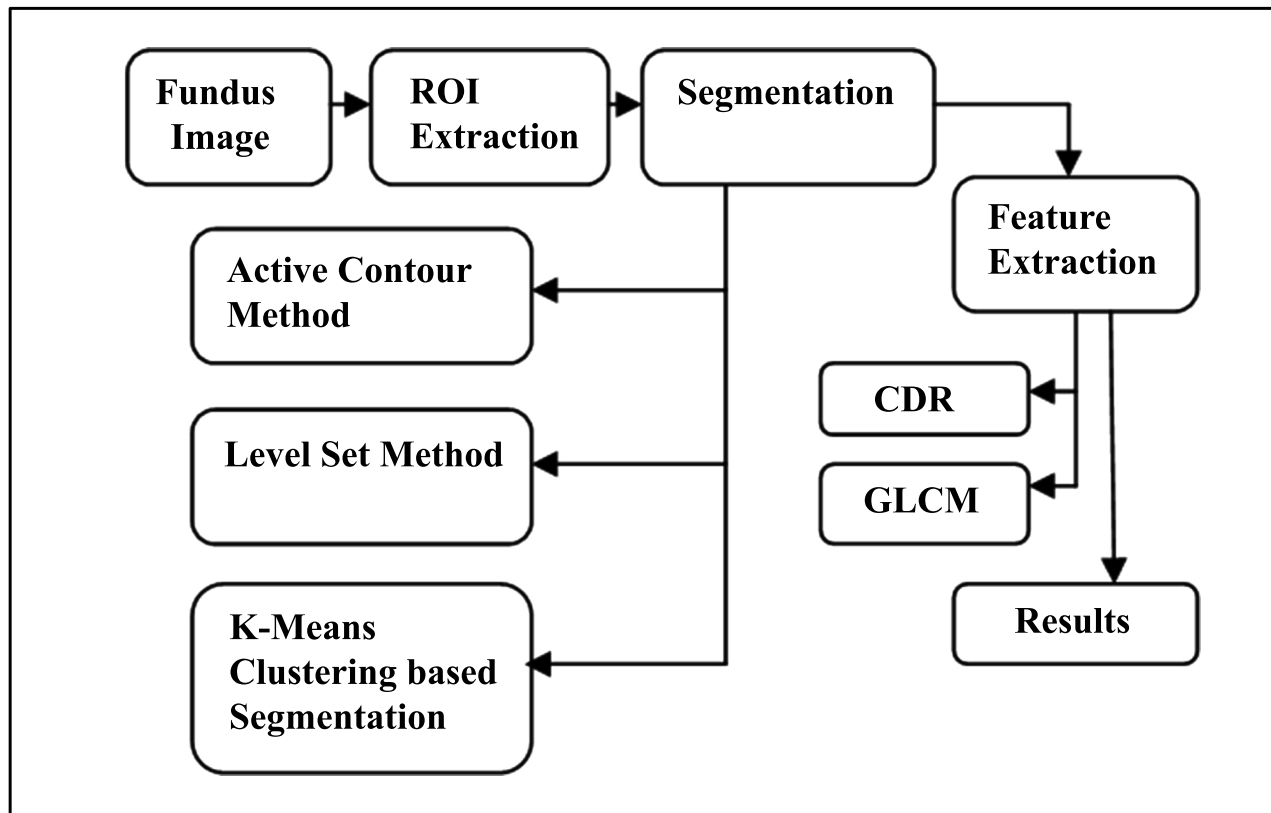


Fig. (8). Block diagram of the proposed method to examine glaucoma on the datasets received from the government hospital database.

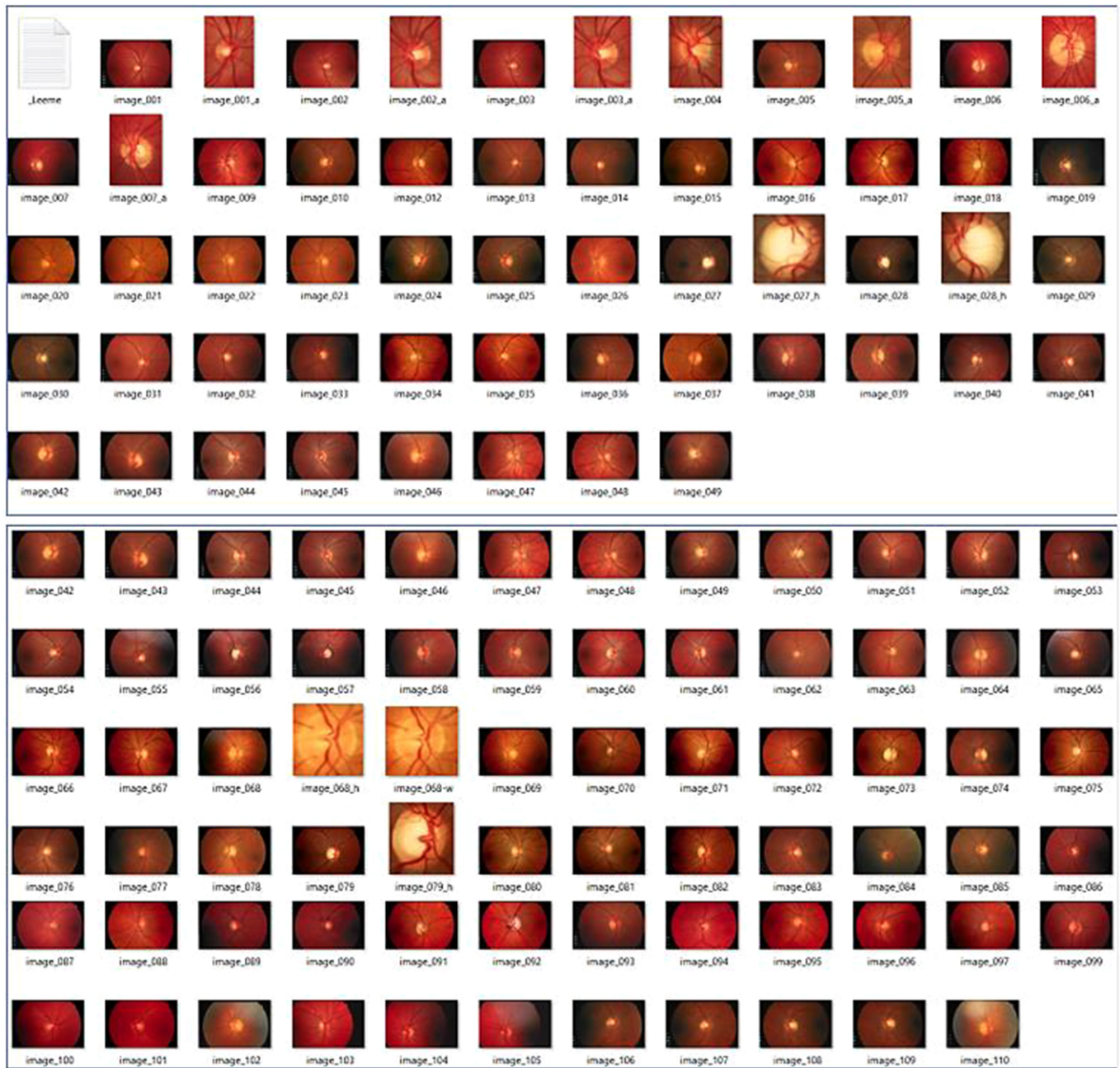


Fig. (9). Undertaken pupil images sample datasets for validation of the proposed design.

5.1.1. Active Contour Segmentation for Optic Disc Localization

The dynamic shape strategies give a successful way to the division of optic circles in which the limits of the plate are distinguished by developing bends. The angle-based dynamic shape model is utilized for plate limit identification.

The calculation holds the shape component of the ideal plate, and its exhibition depends on the form statement. The fundamental point of the strategy is to limit the

energy capability displayed in condition (1), as presented in Eq. (11). The underlying form focuses on physically restricting the limit of optic plate locale. The optic circle locale is sectioned from the pre-handled dark scale picture displayed in Fig. (11). ChanVese (CV) model is additionally a functioning form model. The model is additionally reclassified to separate the OD locale from the comparable trademark districts around it by coordinating data from the various pictures, including channels.

$$E_{snake} = E_{internal} + E_{external} + E_{constraint} \quad (11)$$

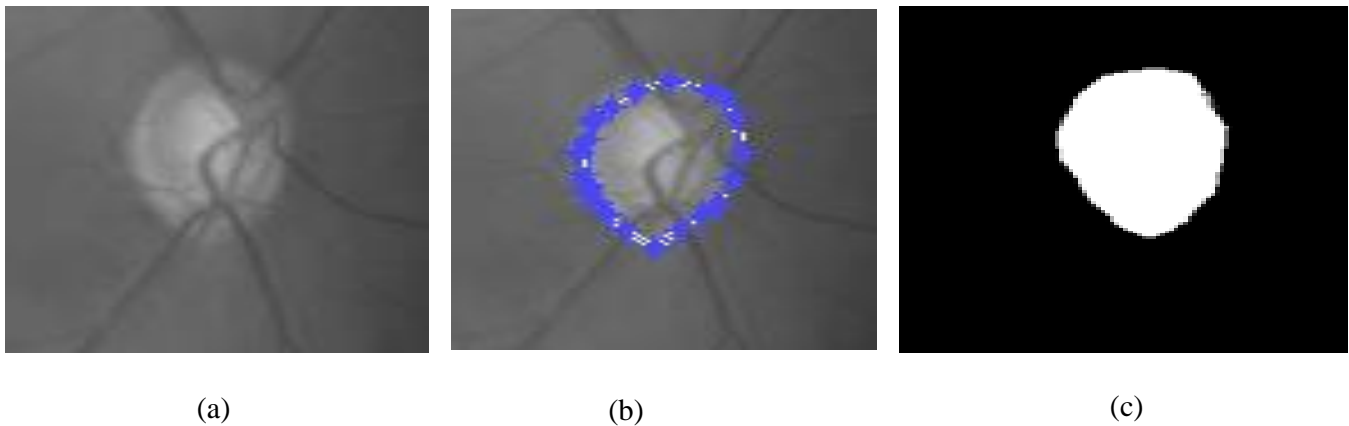


Fig. (10). Results of active contour technique, (a) Original fundus image, (b) Initialisation of contour, (c) Segmented optic disc.

5.1.2. Level Set Algorithm for Optic Cup Localisation

Level set techniques are utilized for performing shape advancement. The capability $\varphi(I, j, t)$ (level set capability) is characterized where (I, j) are arranged in the picture plane and t addresses 'time'. At any coordinated time, the level set capability simultaneously characterizes an edge form and a division of the picture. The edge form is taken to be the zero-level set $\{(i, j) \text{ s. t. } \varphi(I, j, t) = 0\}$, and the division is by the two districts $\{\varphi \geq 0\}$ and $\{\varphi < 0\}$ [48]. The level set capability will be advanced by some incomplete differential condition. If the forefront is characterized to be in the locale where $\varphi < 0$, then the foundation found by this division would be the area inside the circle. Fig. (12) shows the results of optic cup segmentation using the level set method.

Global thresholding function,
$$g(x, y) = \begin{cases} 0 & f(x, y) < T \\ 1 & f(x, y) \geq T \end{cases} \quad (12)$$

Where, $g(x,y)$ - processed image, $f(x,y)$ - Grayscale image, T - average threshold value

The user can select a single global threshold for a picture or different thresholds for each of the 2D slices of the image using global thresholding as given in Equation 11. The binormal fit can be applied to the two-peak histogram and a threshold can be set at the inner peak minimum, as indicated by the normal fits.

Algorithm 1: K-means algorithm with Euclidean distance

```

Input: K, unclassified data sets  $\{X_1, X_2 \dots X_N\}$ 
Output: cluster center  $C_k$  and the consignment of the data points  $\{S_{nk}\}$ .

Randomly initialize  $\{C_k\}$ 
Repeat
For n: =1 to N do
    For k: =1 to K do
        If  $k = \arg \min_i \|C_i - X_n\|^2$ , then  $S_{nk} = 1$ 
        Else  $S_{nk} = 0$ 
    End
End
For k: =1 to K do
 $C_k := \frac{\sum_{n=1}^N X_n S_{nk}}{\sum_{n=1}^N S_{nk}}$ 
End
Until  $\{C_k\}$  or  $\{S_{nk}\}$  don't change.
    
```

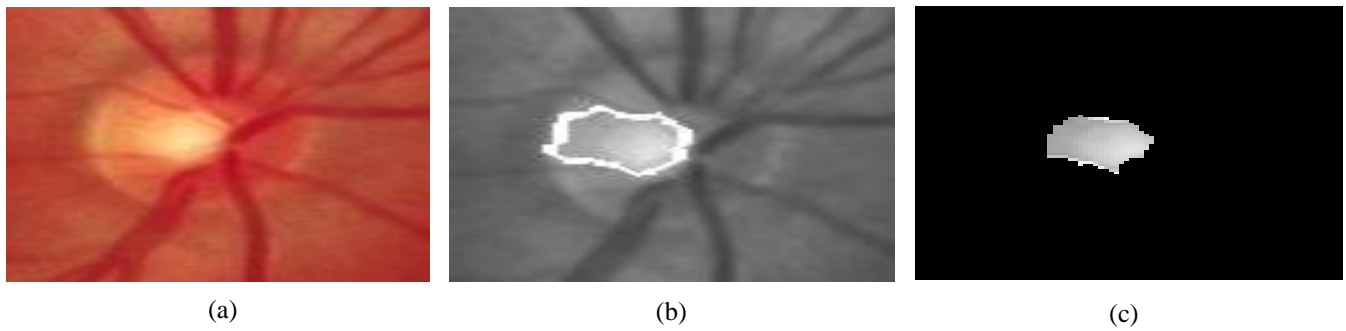


Fig. (11). Results of level set technique, (a) Original fundus image, (b) Cup boundary, (c) Segmented cup region.

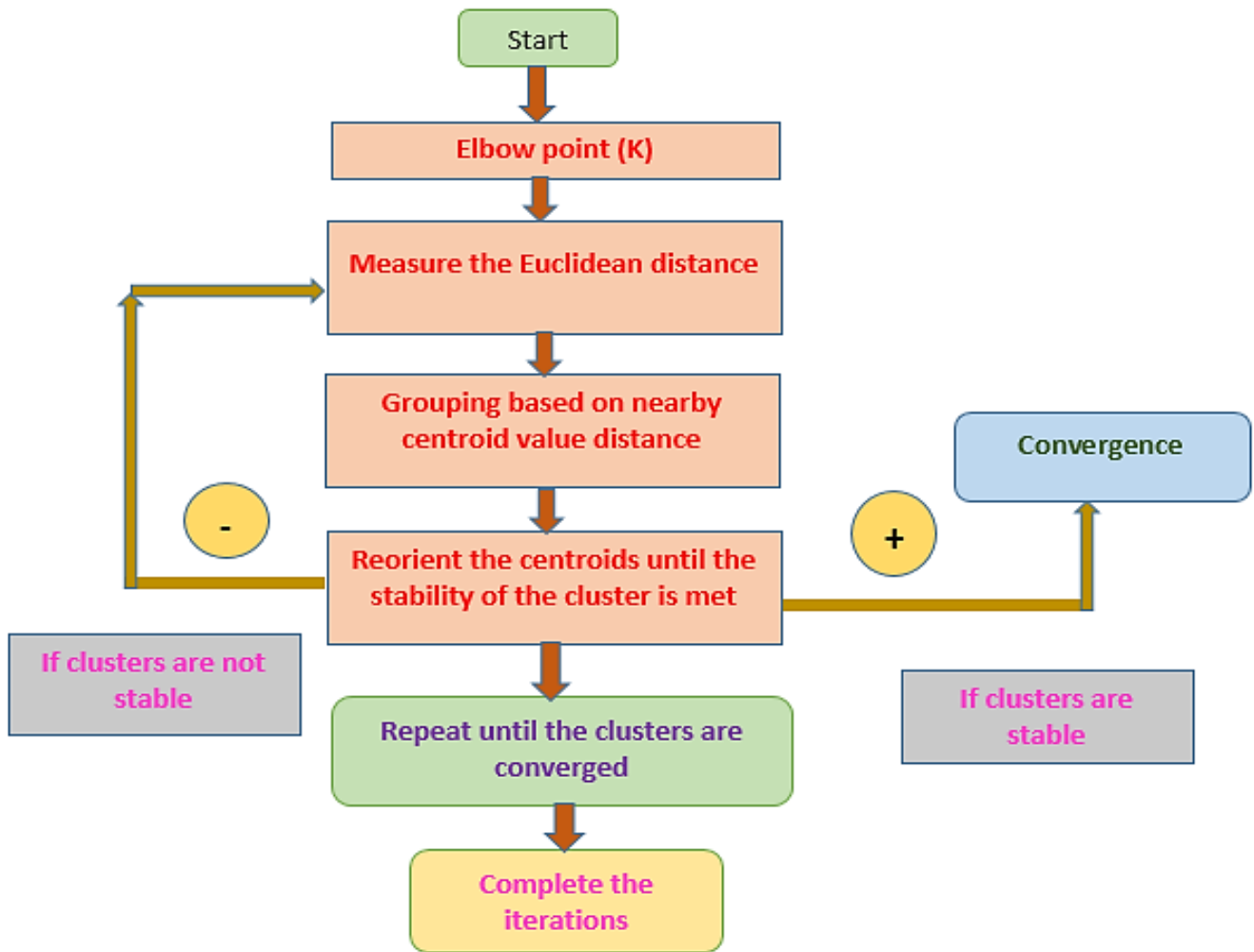


Fig. (12). K-means algorithm for the applied datasets on the current work.

Several experimental solutions have also been offered to allow automatic threshold choosing. Equation 12 explains the global thresholding function. The segmented image is output slices and in packed bit (0,1) format using the thresholding option. Pixels with intensities below the

threshold are all set to 0, while the remaining pixels are all set to 1.

In k means (delineated in Algorithm 1), primarily k initial points are preferred where k is a constraint describing the number of clusters to be required and

which is to be demarcated at the start. The flowchart of the K-means clustering algorithm execution procedure is summarized as follows in Fig. (13). The K-Means clustering algorithm is used for clustering the fundus image into different clusters, with each cluster of different properties. This method of group evaluation aims to divide 'n' perceptions into k bunches so that each perception belongs to the group with the closest mean. The fundus image in the suggested work is divided into three groups: red, blue, and green. The ROI fundus picture is divided into K clusters using an iterative method called the K-means algorithm in this suggested study. The optic disc, optic cup, and a small amount of additional image areas are mostly covered by the ROI. The segmentation of the optic disc and optic cup is completed following the clustering procedure. The grey level of each pixel is compared to a single threshold picture using the thresholding technique. Choosing the threshold value, the region of the optic disc and cup is segmented separately. The objects are isolated by converting grayscale images into binary images.

To validate the AUTO-ML characteristics and its performance efficiency, glaucoma datasets are undertaken. To run AutoML on specified datasets, the Auto Weka 2.0 package manager is enabled in the existing WEKA software after the experimental run is complete on the given glaucoma dataset. The mainstream pipeline streamlining agent was initially evolved to computerize

the biomedical information pool, yet it was made path-actualized to deal with advanced AI tasks [49]. This stage is considered the initial stage of the Tree-based Pipeline Optimization Tool (TPOT) and prevails as the open-source hereditary program platform-based AutoML framework that is implied to deal with initial data preprocessing, model choice, and meta parameter enhancement undertakings on the AI issue. TPOT analysis with the scikit-learn platform incorporated with the Python AI library package initiates the administrator to compare an AI calculation that is available in that library. In observational analyses, the Auto-sklearn framework accomplished well than or similarly to the functioning of the Auto-WEKA framework to the scale of 90% and likewise the best in front of the pack in the recently referenced AutoML challenge.

Notwithstanding its great execution, the first Auto-sklearn stage is restricted to taking care of datasets of generally humble size and stretched out to deal with bigger datasets. The freshest arrival of the framework, known as PoSH (PORTfolio Successive Halving) Auto-sklearn, utilizes the BOHB calculation depicted in segment II and has made a huge expansion and speed rate associated with the Auto-sklearn framework. The proposed Auto-sklearn framework attained better results in the second cycle of the AutoML challenge. In any case, Auto-sklearn, along with Auto-WEKA platforms are as yet no exceptions to deal with huge quantifiable healthcare datasets.

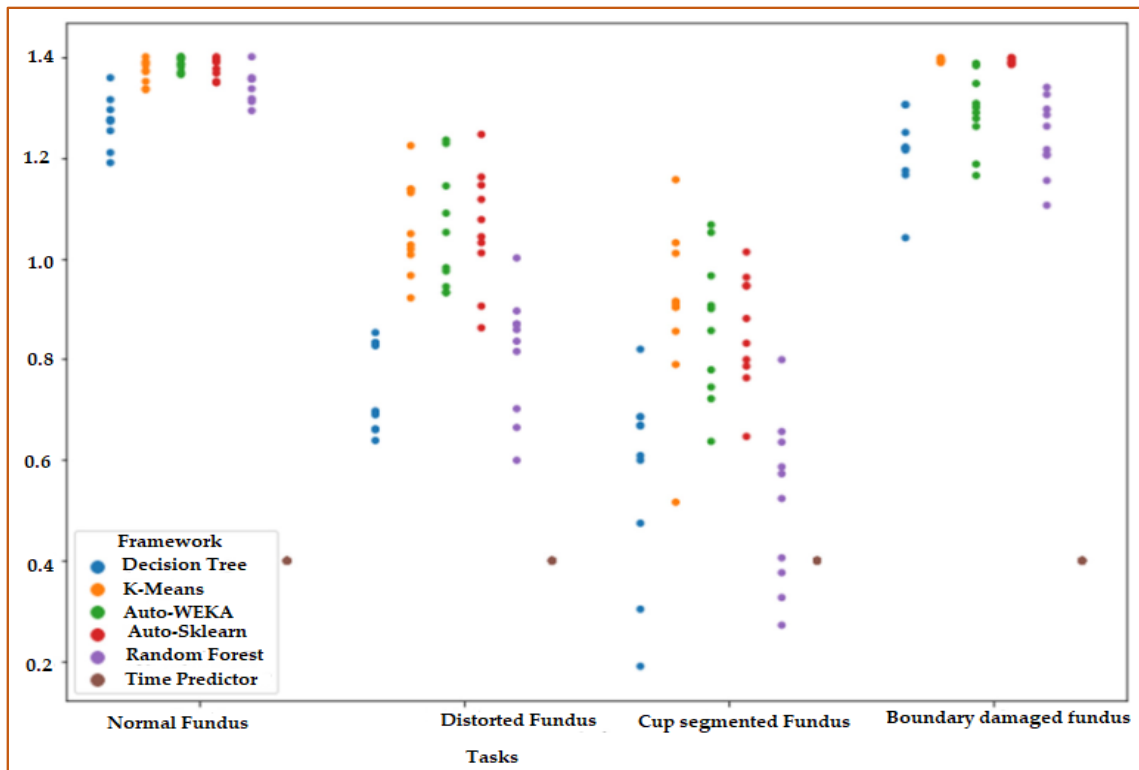


Fig. (13). Comparison of AutoML with existing AI algorithms.

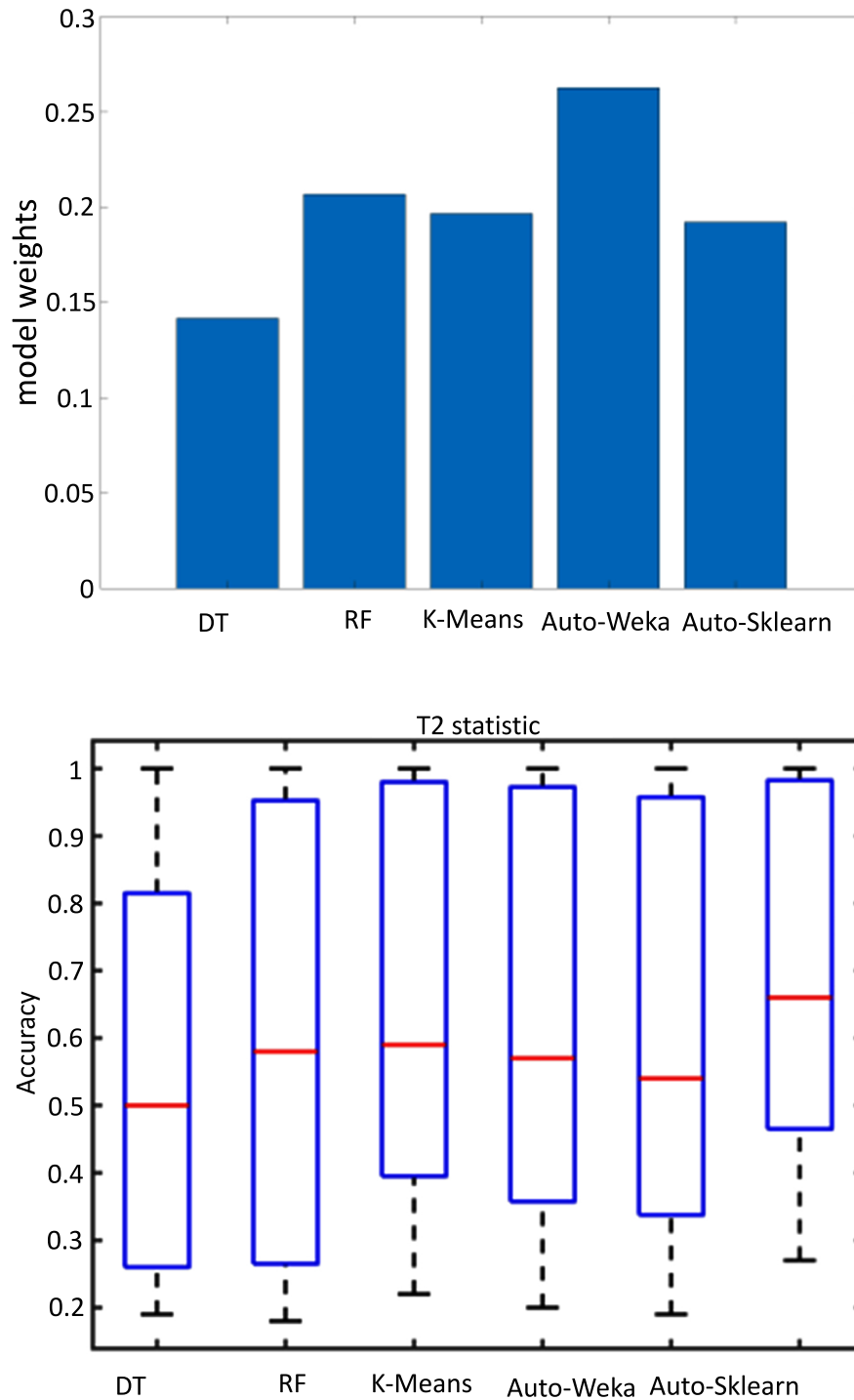


Fig. (14). Model analysis using model weightage and T2 statistics.

With a total computing duration of 120 minutes, the comparison test was run using a timeframe of 2 minutes. The visual depiction of the results is shown in Fig. (14). The datasets are on the X-axis, and the AUROC score is displayed on the Y-axis. Each of the ten-fold scores for an AutoML approach is shown by a colored dot. Fig. (14)

shows that the decision tree and Random Forest do not always beat the time predictor on the common fundus dataset. For the distorted fundus dataset, Auto-Sklearn falls behind the other three algorithms, but it outperforms the time predictor and is on par with K-Means in terms of performance. All AutoML algorithms occupy the maximum

score on at least a one-fold basis on the cup-segmented fundus dataset. All methods perform well on the boundary-damaged fundus data set, given their median scores and distribution [50].

Model evaluation is performed through Auto-Weka analysis following model selection. For model weightage analysis, additional criteria, including Decision Tree, Random Forest, K-means, and Auto-Sklearn, are used. Fig. (14) presents the model assessment outcomes for two monitoring statistics that were derived by averaging the performance outcomes based on the boundary-damaged fundus datasets. As can be observed, Auto-Weka and RF outperform Auto-Sklearn and Decision Tree, demonstrating that these models are more effective in extracting nonlinear, non-Gaussian, and dynamic information. Additionally, the Auto-Weka of the suggested framework for diagnosing fundus abnormalities performed amazingly under the T2 statistic.

CONCLUSION

AutoML is a major thrust area of software engineering that can help non-specialists use AI in the Electronics-health decision-making system. The proposed work elaborates on all the configurations of the e-healthcare system design and development. Further, to perform proactive decision-making by the clinical experts, the cloud-retrieved data samples are analyzed, and determined results are attained by using AutoML implementation. Also, the AutoML execution platform AutoWeka 2.0 package and its entire optimization with feature extractions pipeline mechanisms are described in a detailed manner. When data is kept in a cloud environment, the suggested method in this study offers authentication and a storage strategy that strengthens user health data. The proposed work affords better security standards for data transmission using a combination of AES-RSA as a cryptographic technique.

Once AutoML standardization has been established, more automation may be employed to speed up processes to identify the best-fit model for its better interpretation with minimum computation time. For example, promising outcomes inferred from the case study of glaucoma demonstrate that clinical content mining can be actualized without the need for enormous named datasets ordinarily important for AI. The experimentation initiated with ROI feature extraction followed by an active level set algorithm for cup localization to identify the fundus images automatically without any human intervention using AutoML within 2 minutes of computation time. At last, the proposed framework can be conceivably utilized by nearly any clinician without extraordinary informatics, which is a restriction of many existing AutoML frameworks. The platform will afford the best classifier with a suitable optimizer automatically by taking high-influencing hyperparameters on their own during the data mining process without any external inputs.

The inferences based on various AutoML features are summarized as follows

- The NLP framework called “Clinical Text Extraction,

Reasoning, and Mapping System” (CTERMS) was later used to recognize twisted data in clinical notes

- The best usage of populace-oriented techniques is the covariance lattice adaption advancement methodology (CLA-AM), which tests the multivariate Gaussian appropriation levels

- BOHB calculation has joined the standards of Bayesian streamlining and Hyperband to integrate the more proficient hyperparameter analyzer

- Auto-WEKA was the main AutoML framework that supports progressive meta parameter streamlining issues by automatic algorithm upgradation

- Auto-Sklearn seems to be a robotized outfit development model evaluated during the enhancement

Limitations and Future Scopes:

- Big data from the healthcare informatics world with understandable biomedical context are inefficient for the machine learning pipeline optimization techniques used currently.

- The goal of AutoML Challenges is to resolve supervised machine learning issues in the absence of human involvement under certain computing limitations.

- Qualities are often created physically by a human professional through the implementation of trial and error since they frequently need in-depth topic expertise.

- Feature engineering has become a vital, time-consuming platform in the machine-learning process.

To increase variety and generalizability, future research might be expanded to include different sites and employ AutoML techniques.

- To improve the efficacy of illness prediction, the text analysis and picture analysis techniques may be expanded to include video content analysis.

- Use of predictive models that can analyze several social media sites at once to produce precise and timely forecasts of outbreaks

- Combining the text content of user postings with input characteristics like emotional content, comments, locations, *etc.*, to improve analysis and prediction of health occurrences and other relevant events

LIST OF ABBREVIATIONS

AutoML = Automated Machine Learning

AI = Artificial Intelligence

CHFS = Closed-loop Healthcare Feature Store

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

HUMAN AND ANIMAL RIGHTS

In the present work, no humans or animals were trailed or tested in real-time laboratory experimentation.

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data supporting the findings of the article is available in the published article result and discussion sections.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the host institution, Engineering College, for affording better financial support and a research platform with highly configured software to perform experimentation analysis.

REFERENCES

- [1] N.D.A. Tarmizi, F. Jamaluddin, A. Abu Bakar, Z.A. Othman, S. Zainudin, and A.R. Hamdan, "Malaysia dengue outbreak detection using data mining models", *J. Next Gener. Inf. Technol.*, vol. 4, pp. 96-107, 2013.
- [2] Y. Chen, J.H. Chang, A.S. Greenlee, K.C. Cheung, A.H. Slocum, and R. Gupta, "Multi-turn, tension-stiffening catheter navigation system", *2010 IEEE International Conference on Robotics and Automation*, 2010pp. 5570-5575
[<http://dx.doi.org/10.1109/ROBOT.2010.5509786>]
- [3] S. Archana, and DR Elangovan, "Survey of classification techniques in data mining", *Int. J. Comput. Sci. Mob. Appl.*, vol. 2, pp. 65-71, 2014.
- [4] S. Vijayarani, and S. Dhayanand, "Liver disease prediction using SVM and naive bayes algorithms", *IJSETR*, vol. 4, pp. 816-820, 2015.
- [5] A. Gulia, R. Vohra, and P. Rani, "Liver patient classification using intelligent techniques", *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, pp. 5110-5115, 2014.
- [6] P. Rajeswari, and G.S. Reena, "Analysis of liver disorder using data mining algorithm", *Glob. J. Comput. Sci. Technol*, vol. 10, pp. 48-52, 2010.
- [7] T.N. Do, T. Tjahjowidodo, M.W.S. Lau, T. Yamamoto, and S.J. Phee, "Hysteresis modeling and position control of tendon-sheath mechanism in flexible endoscopic systems", *Mechatronics*, vol. 24, no. 1, pp. 12-22, 2014.
[<http://dx.doi.org/10.1016/j.mechatronics.2013.11.003>]
- [8] N.G. Cheng, M.B. Lobovsky, S.J. Keating, A.M. Setapen, K.I. Gero, A.E. Hosoi, and K.D. Iagnemma, "Design and analysis of a robust, low-cost, highly articulated manipulator enabled by jamming of granular media. In2012", *IEEE Int. Conf. Robot. Autom.*, pp. 4328-4333, 2012.
- [9] F. Ibrahim, M.N. Taib, W.A.B.W. Abas, C.C. Guan, and S. Sulaiman, "A novel dengue fever (DF) and dengue haemorrhagic fever (DHF) analysis using artificial neural network (ANN)", *Comput. Methods Programs Biomed.*, vol. 79, no. 3, pp. 273-281, 2005.
[<http://dx.doi.org/10.1016/j.cmpb.2005.04.002>] [PMID: 15925426]
- [10] A.M. Rahmani, E. Yousefpoor, M.S. Yousefpoor, Z. Mehmood, A. Haider, M. Hosseinzadeh, and R. Ali Naqvi, "Machine learning (ML) in medicine: Review, applications, and challenges", *Mathematics*, vol. 9, no. 22, p. 2970, 2021.
[<http://dx.doi.org/10.3390/math9222970>]
- [11] F.M. Ba-Alwi, and H.M. Hintaya, "Comparative study for analysis the prognostic in hepatitis data: Data mining approach", *Int. J. Sci. Eng. Res.*, vol. 4, pp. 680-685, 2013.
- [12] B. Karlik, "Hepatitis disease diagnosis using back propagation and the naive bayes classifiers", *J. Sci. Technol.*, vol. 1, pp. 49-62, 2011.
- [13] G. Sathyadevi, "Application of CART algorithm in hepatitis disease diagnosis", *2011 International Conference on Recent Trends in Information Technology (ICRTIT)*, 2011 Chennai, India
[<http://dx.doi.org/10.1109/ICRTIT.2011.5972349>]
- [14] Y. Singh, P.K. Bhatia, and O. Sangwan, "A review of studies on machine learning techniques", *IJCSS*, vol. 1, pp. 70-84, 2007.
- [15] T.P. Chenal, J.C. Case, J. Paik, and R.K. Kramer, "Variable stiffness fabrics with embedded shape memory materials for wearable applications", *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2014pp. 2827-2831
[<http://dx.doi.org/10.1109/ROBOT.2014.6942950>]
- [16] C. Chakraborty, and A. Abougreen, "Intelligent internet of things and advanced machine learning techniques for covid-19", *EAI Endorsed Trans. Pervasive Health Technol.*, vol. 7, no. 26, 2021.
- [17] A. Kishor, C. Chakraborty, and W. Jeberson, "Intelligent healthcare data segregation using fog computing with internet of things and machine learning", *Int. J. Eng. Syst. Model. Simul.*, vol. 12, no. 2/3, pp. 188-194, 2021.
[<http://dx.doi.org/10.1504/IJESMS.2021.115533>]
- [18] C. Chakraborty, and A. Kishor, "Real-time cloud-based patient-centric monitoring using computational health systems", *IEEE Trans. Comput. Soc. Syst.*, vol. 9, no. 6, pp. 1613-1623, 2022.
[<http://dx.doi.org/10.1109/TCSS.2022.3170375>]
- [19] Y. Chen, R. Xia, K. Yang, and K. Zou, "DGCA: high resolution image inpainting via DR-GAN and contextual attention", *Multimedia Tools Appl.*, vol. 82, no. 30, pp. 47751-47771, 2023.
[<http://dx.doi.org/10.1007/s11042-023-15313-0>]
- [20] E.B. Priyanka, S. Thangavel, X.Z. Gao, and N.S. Sivakumar, "Digital twin for oil pipeline risk estimation using prognostic and machine learning techniques", *J. Ind. Inf. Integr.*, vol. 26, p. 100272, 2022.
[<http://dx.doi.org/10.1016/j.jii.2021.100272>]
- [21] E.B. Priyanka, S. Thangavel, X.Z. Gao, and N.S. Sivakumar, "IoT world communication through antenna propagation with emerging design analysis features", In: *Next-Generation Antennas: Advances and Challenges*, Wiley, 2021.
[<http://dx.doi.org/10.1002/9781119792765.ch11>]
- [22] E.B. Priyanka, S. Thangavel, G. Manikandan, and C.V. Rahav, "Artificial intelligence impact on pattern classification in association with IoT for advanced applications", In: *Cyber Security and Operations Management for Industry 4.0*, 1st ed. CRC Press, 2022.
[<http://dx.doi.org/10.1201/9781003212201-5>]
- [23] V. Hassani, T. Tjahjowidodo, and T.N. Do, "A survey on hysteresis modeling, identification and control", *Mech. Syst. Signal Process.*, vol. 49, no. 1-2, pp. 209-233, 2014.
[<http://dx.doi.org/10.1016/j.ymssp.2014.04.012>]
- [24] B.E. Schubert, and D. Floreano, "Variable stiffness material based on rigid low-melting-point-alloy microstructures embedded in soft poly(dimethylsiloxane) (PDMS)", *RSC Adv.*, vol. 3, no. 46, pp. 24671-24679, 2013.
[<http://dx.doi.org/10.1039/c3ra44412k>]
- [25] X. He, K. Zhao, and X. Chu, "AutoML: A survey of the state-of-the-art", *Knowl. Base. Syst.*, vol. 212, p. 106622, 2021.
[<http://dx.doi.org/10.1016/j.knsys.2020.106622>]
- [26] P. Gijsbers, E. LeDell, J. Thomas, S. Poirier, B. Bischl, and J. Vanschoren, "An open source AutoML benchmark", *arXiv:1907.00909*, 2019.
- [27] I. Guyon, L. Sun-Hosoya, M. Boullé, H.J. Escalante, S. Escalera, Z. Liu, D. Jajetic, B. Ray, M. Saeed, M. Sebag, and A. Statnikov, "Analysis of the automl challenge series", *AutoML*, vol. 20, no. 19, pp. 177-190, 2019.
- [28] Y. He, J. Lin, Z. Liu, H. Wang, L.J. Li, and S. Han, "Amc: Automl for model compression and acceleration on mobile devices", *Proceedings of the European conference on computer vision*

- (ECCV), 2018pp. 784-800
- [29] E.B. Priyanka, S. Thangavel, and P. Prabhakaran, "Rank-based risk target data analysis using digital twin on oil pipeline network based on manifold learning", *Proc. Inst. Mech. Eng. E: J. Process Mech. Eng.*, vol. 236, no. 4, 2022.
[<http://dx.doi.org/10.1177/09544089211073241>]
- [30] E.B. Priyanka, and S. Thangavel, "Multi-type feature extraction and classification of leakage in oil pipeline network using digital twin technology", *J. Ambient Intell. Humaniz. Comput.*, vol. 13, no. 12, pp. 5885-5901, 2022.
[<http://dx.doi.org/10.1007/s12652-022-03818-9>]
- [31] M. Ponnibala, E.B. Priyanka, and S. Thangavel, "Proliferative diabetic retinopathy diagnostic investigation using retinal blood vessels mining technique", *Sens. Imaging*, vol. 22, no. 1, p. 10, 2021.
[<http://dx.doi.org/10.1007/s11220-021-00331-9>]
- [32] I. Guyon, I. Chaabane, H.J. Escalante, S. Escalera, D. Jajetic, J.R. Lloyd, N. Macià, B. Ray, L. Romaszko, M. Sebag, and A. Statnikov, "A brief review of the ChaLearn AutoML challenge: Any-time any-dataset learning without human intervention", In: *Workshop on Automatic Machine Learning*, PMLR, 2016, pp. 21-30.
- [33] E. Real, C. Liang, D. So, and Q. Le, "Automl-zero: Evolving machine learning algorithms from scratch", *Proceedings of the 37th International Conference on Machine Learning, Online*, PMLR 119, 2020
- [34] A. Truong, A. Walters, J. Goodsitt, K. Hines, C.B. Bruss, and R. Farivar, "Towards automated machine learning: Evaluation and comparison of AutoML approaches and tools", *2019 IEEE 31st international conference on tools with artificial intelligence ICTAI*, 2019pp. 1471-1479
- [35] M. Wever, A. Tornede, F. Mohr, and E. Hüllermeier, "AutoML for multi-label classification: Overview and empirical evaluation", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 9, pp. 3037-3054, 2021.
[<http://dx.doi.org/10.1109/TPAMI.2021.3051276>] [PMID: 33439834]
- [36] J. Yang, R. Shi, and B. Ni, "Medmnist classification decathlon: A lightweight automl benchmark for medical image analysis", *arXiv:2010.14925*, 2021.
[<http://dx.doi.org/10.1109/ISBI48211.2021.9434062>]
- [37] M.K. Shende, S.Q. Salih, N.D. Bokde, M. Scholz, A.Y. Oudah, and Z.M. Yaseen, "Natural time series parameters forecasting: Validation of the pattern-sequence-based forecasting (PSF) algorithm; A new python package", *Appl. Sci.*, vol. 12, no. 12, p. 6194, 2022.
[<http://dx.doi.org/10.3390/app12126194>]
- [38] S. Guan, B. Zhao, Z. Dong, M. Gao, and Z. He, "GTAD: Graph and temporal neural network for multivariate time series anomaly detection", *Entropy*, vol. 24, no. 6, p. 759, 2022.
[<http://dx.doi.org/10.3390/e24060759>] [PMID: 35741480]
- [39] S. Chatterjee, R. Bopardikar, M. Guerard, U. Thakore, and X. Jiang, "MOSPAT: AutoML based model selection and parameter tuning for time series anomaly detection", *arXiv:2205.11755*, 2022.
- [40] K.W. Wan, C.H. Wong, H.F. Ip, D. Fan, P.L. Yuen, H.Y. Fong, and M. Ying, "Evaluation of the performance of traditional machine learning algorithms, convolutional neural network and AutoML Vision in ultrasound breast lesions classification: A comparative study", *Quant. Imaging Med. Surg.*, vol. 11, no. 4, pp. 1381-1393, 2021.
[<http://dx.doi.org/10.21037/qims-20-922>] [PMID: 33816176]
- [41] Y. Chen, R. Xia, K. Zou, and K. Yang, "FFTI: Image inpainting algorithm via features fusion and two-steps inpainting", *J. Vis. Commun. Image Represent.*, vol. 91, p. 103776, 2023.
[<http://dx.doi.org/10.1016/j.jvcir.2023.103776>]
- [42] J. Zhang, H. Huang, X. Jin, L.D. Kuang, and J. Zhang, "Siamese visual tracking based on criss-cross attention and improved head network", *Multimedia Tools Appl.*, vol. 83, pp. 1589-1615, 2023.
- [43] J. Zhang, X. Zou, L.D. Kuang, J. Wang, R.S. Sherratt, and X. Yu, *CCTSDB 2021: A more comprehensive traffic sign detection benchmark*, University of Reading, 2022.
- [44] J. Zhang, Z. Zheng, X. Xie, Y. Gui, and G.J. Kim, "ReYOLO: A traffic sign detector based on network reparameterization and features adaptive weighting", In: *JAISE*, vol. 14. 2022, no. 4, pp. 1-18.
- [45] J. Waring, C. Lindvall, and R. Umeton, "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare", *Artif. Intell. Med.*, vol. 104, p. 101822, 2020.
[<http://dx.doi.org/10.1016/j.artmed.2020.101822>] [PMID: 32499001]
- [46] F. Sabry, T. Eltaras, W. Labda, K. Alzoubi, and Q. Malluhi, "Machine learning for healthcare wearable devices: The big picture", *J. Healthc. Eng.*, vol. 2022, pp. 1-25, 2022.
[<http://dx.doi.org/10.1155/2022/4653923>] [PMID: 35480146]
- [47] P. Esmailzadeh, "Use of AI-based tools for healthcare purposes: A survey study from consumers' perspectives", *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 1, p. 170, 2020.
[<http://dx.doi.org/10.1186/s12911-020-01191-1>] [PMID: 32698869]
- [48] M. Javaid, A. Haleem, R. Pratap Singh, R. Suman, and S. Rab, "Significance of machine learning in healthcare: Features, pillars and applications", *IJIN*, vol. 3, pp. 58-73, 2022.
[<http://dx.doi.org/10.1016/j.ijin.2022.05.002>]
- [49] A.U. Haq, J.P. Li, A. Saboor, J. Khan, S. Wali, S. Ahmad, A. Ali, G.A. Khan, and W. Zhou, "Detection of breast cancer through clinical data using supervised and unsupervised feature selection techniques", *IEEE Access*, vol. 9, pp. 22090-22105, 2021.
[<http://dx.doi.org/10.1109/ACCESS.2021.3055806>]
- [50] R. Poonguzhali, S. Ahmad, S.P. Thiruvannamalai, B.S. Anantha, P. Joshi, J.G. Prasad, and S. Won Kim, "Automated brain tumor diagnosis using deep residual u-net segmentation model", *Comput. Mater. Continua*, vol. 74, no. 1, pp. 2179-2194, 2023.
[<http://dx.doi.org/10.32604/cmc.2023.032816>]