

A Context-Aware M-RIPPER Algorithm for Heart Disease Prediction

Saad Almutairi

University of Tabuk

Manimurugan S (✉ mmurugan@ut.edu.sa)

University of Tabuk <https://orcid.org/0000-0003-1837-6797>

Majed Mohammed Aborokbah

University of Tabuk

Narmatha C

University of Tabuk

Subramaniam Ganesan

Oakland University

Riyadh A AlzahebHani

University of Tabuk

Hani Almoamari

Islamic University of Madinah

Research Article

Keywords: Context Awareness, MRIPPER, IoT, Cloud Computing, Rule based learning, heart disease

Posted Date: May 4th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1545620/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Journal of Healthcare Engineering on July 11th, 2022. See the published version at <https://doi.org/10.1155/2022/7853604>.

Abstract

These days, mobile computing devices are ubiquitous and widely used in almost every facet of daily life. Computing and the modern world are not really coexisting anymore. With a wide range of conditions and areas of concern, the medical domain was also concerned. New types of technologies, such as context-aware systems and applications, are constantly being infused into the medicine field. An IoT-enabled healthcare system based on context awareness is developed in this work. In order to collect and store patient data, smart medical devices are employed. Context-aware data from the database includes the patient's medical records and personal information. The MRIPPER (Modified Repeated Incremental Pruning to Produce ErrR) technique is used to analyze and classify the data. A rule-based machine learning method is used in this algorithm. The rules for analyzing data sets in order to make predictions about heart disease are framed using this algorithm. MATLAB is used to simulate the proposed model's performance analysis. Other models like random forest, J48, CART, JRip, and OneR algorithms are also compared to validate the proposed model's performance. The proposed model obtains 98.89 percent accuracy, 96.76 percent precision, 99.05 percent sensitivity, 94.35 percent specificity, and 97.60 percent f-score. Predictions for subjects in the normal and abnormal classes were both accurate with 97.38 for normal and 97.93 for abnormal.

1. Introduction

Context Awareness originated as a term from Ubiquitous computing, which is turning into a reality that highlights the integration between the data space and the physical space. With its assistance, individuals could receive and process data whenever and anywhere through a device that can link any internet. Therefore, it can lessen the difficulty of utilizing the device and make individuals' lives simpler and progressively effective. The environment of user in ubiquitous computing, for example, the location, or terminal equipment, and so on is continually changing, which is called context. As part of the central zones of ubiquitous computing, context aware computing has become increasingly very well known among people. Numerous authors have described context according to their comprehension with an exertion to review an extensive basic idea of the context [1]. Schilit and Theimer utilized the term context aware in 1994 and refer to context as location, identities of objects and nearby people and changes to those objects. The term context has been sorted into two classes (logical and physical) [2]. Physical context could be decided through hardware sensor and logical context was either provided through the user's feedback or by observing their communications with the services accessible. For instance, by monitoring or reviewing the user's profile, working schedules, activities, composing movement, and so on. Most research around there utilizes physical sensor for movement, sound, touch, temperature, light, and of course location. The logical sensor, though give associated data by reading user's data from public website pages and different archives and furthermore reviews user's information (interaction) and dependent on those interaction target publicizing [2].

A context can be various elements or factors such as Location, User Identity, Time, Activity, Current Task, Environment, and Hardware. Context-awareness refers that one can utilize context data. A system was context-aware if that it could extricate, decrypt and use context data and modify their performance to the present context of usage. The name context-aware computing was generally perceived through those performing in context-aware, where it was considered that context was a source in its attempt to distribute

and directly combine computer advancement into our lives. Context-aware systems can modify their activities dependent on the present context. This likewise expands adequacy by considering environmental context. CAS observes the condition constantly and proposes reasonable recommendations to users which they could make important actions [3].

The context-aware healthcare system helps hospitals to enhance performance efficiency by incorporating real-time contextual data into the actual workflow like the locations and current conditions of medical devices and employees. It also allows access to environmental data in order to deliver the best possible patient experiences. The solution supports both real-time asset monitoring and event-driven tracking, with real-time tracking moving across the surrounding and event-driven tracking existing and entering regions. The primary goal of health care systems must be to safeguard and maintain the security of patient data. In addition, the use of contextual is critical in interactive frameworks where the user's data changes frequently, such as in portable and ubiquitous computing (Fig. 1).

1.1. CONTEXT AWARE IN HEALTHCARE

The combination of Internet and the Medical domain is a progressive innovation, with present research aiming on the use of computing to support in training among the medical sector. The smart clinical devices market was predicted to reach above 126 billion dollars profit by the year 2028 while smart wearable devices intended to be extensively utilized to accomplish enhanced health, quality of life, and protection of people. Moreover, to their capability to aid real-time constant observance of patient's data, such devices also make context-aware mobility significance to enhance overall condition of medical care. Context Aware System (CAS) is a system that can adjust their activities to context changes without unequivocal user intercession. The CAS platform should unequivocally present by its component's functionalities, context data, and the control activity and furthermore provides services to users utilizing context data where pertinence relies upon the user's operation. In this way, a context aware domain could be intended middleware support that permits the exchange of environmental data out of the minimum infrastructure range to a more significant range for definition and decision. This multi-layered design was common for the Cloud computing sequence that permits setting the middleware layer as a major aspect of a Sensor-Cloud interface in the layer of PaaS (Platform as a Service) [4].

Context aware systems also play an important role in health care systems, whereby automatically distinguishing a patient from the rest of the surroundings, recording the various events associated with a specific patient, keeping track of the various services provided in a specific location, and providing the necessary documentation are some of the important functionalities that must be encountered by the system. There is also the extra obligation on the system to keep the patient's and health care professional's information secure and safe. Security is also required for any equipment used by health care institutions. As a result, health care systems cannot be viewed as a separate system from the rest of the technological systems; rather, they are a sociotechnical system that is dependent on the collaborated results of the communications among the technology and user. Context awareness aids in the more precise diagnoses of the observed patient's health problems. It can recognize behavioural patterns and so make more exact conclusions about people and their surroundings. Adaptation, personalization, and proactivity are the three most essential advantages of context awareness. The following was a breakdown of the three advantages:

Adaptation was focused with tailoring the services or information to the user's present situation. A specific example is when the system in issue adjusts the data it delivers based on networks and device contexts, like speed of connection and resolution of display. Personalization is the process of customizing the system to individual users, where each user sees the framework differently. Personalization was according to the individual user's choices, habit, abilities, duties, and so on. The data or, more accurately, the degree of detail at which information was supplied to doctors, i.e., in Health care Monitoring System, was plainly different from that received to patients or caretakers.

Proactivity was considered with providing services for the users based on forecasts of future circumstances. When it comes to Healthcare Monitoring Systems, proactivity is critical in producing really useful and promising solutions. There are several examples and user case. For example, being proactive and determining health issue situations aids in the discovery of these concerns at the early stages, which frequently enhances the likelihood of averting or, at the very least, minimizing the harm caused by the health issues. Another situation or user case in which proactivity aids in the prediction of diseases-causing mutation induced by genetical alterations in the genomes that were likely to get molecular effects. From the most recent decade, the CAS targets around web applications, and desktop computing to the Internet of Things (IoT). Because of advance sensor innovations, sensors are getting stronger, less expensive and less in size. In this present world, there are numerous sensors and eventually, these sensors create a lot of information, for example, big data. Except if we dissect, interpret and comprehend the information which collected that information may not produce important data. Context-aware computing plays a significant part in handling this task, for example, mobile and pervasive, which would be effective in the IoT model also. This enables us to save the context data associated to sensor information, so the interpretation should be possible all the more effectively, genuinely and furthermore context makes it simpler to execute machine-to-machine interaction as it is the core component in the IoT condition [5].

Heart disease is one of the most critical and difficult health problems in the modern world. Heart disease reduces blood vessel function and causes coronary artery infections, both of which weaken the patient's body, especially in adults and the elderly. According to the WHO, heart diseases are the leading cause of death globally. In 2019, an estimated 17.9 million people died from heart diseases, accounting for 32% of total worldwide mortality. Stroke or heart attack caused 85 percent of these fatalities. More than three-quarters of all heart disease deaths occur in low- and middle-income countries. Heart diseases are responsible for 38% of the 17 million premature deaths (before the age of 70) caused by noncommunicable disease in 2019. Most heart disease can be prevented by addressing behavioural risk factors such as cigarette use, poor diet and obesity, inactivity, and excessive alcohol use. It is critical to detect heart disease as soon as possible so that therapy with counselling and medicines may begin. Heart diseases are a kind of heart and blood vessel disease. Among these include cerebrovascular disease, coronary heart diseases, peripheral artery diseases, congenital heart diseases, rheumatic heart diseases, deep vein thrombosis, and pulmonary embolism [<https://bit.ly/35qpAGG>].

Context awareness performs a significant role in the concept of Internet of Things, as it provides rich contextual knowledge that can make the system perform more effectively. Since every context of healthcare is different, it is important to determine an adequate context aware architecture for IoT healthcare

applications. In this work, a context aware health care method based on the application of IoT was proposed. Smart medical devices were utilized to collect and retain patient data, which was stored in a database. The database contained context-aware data, such as the names, addresses, and medical histories of the patients. A rule-based machine learning technique, a modified RIPPER algorithm, was utilized to analyze and classify the data. The rules for analyzing data for the prediction of heart disease were developed using this algorithm. The remaining part of this research is presented in following sections as, section II discusses the related works, section III presents the proposed methodology, section IV presents the performance analysis, and section V presents the conclusion and future extension of the research.

2. Related Works

Yousef A presented an analysis of healthcare monitoring framework and its offerings on the IoT platforms. Many functions that exist in healthcare systems have been described and modelled. In addition, this work aimed to establish and proposed the general frameworks for the development and design of contexts-aware health care monitoring framework in IoT domain. The essential elements of healthcare monitoring framework, as well as their relationships, were discovered and modelled in such a model. The work also emphasized the importance of the AI sectors in tackling robust context aware healthcare monitoring. This framework was built on a distributed layer architecture, with distinct components implemented across the physical layers, cloud platform, and fog platform [1].

Mohamed A B et al. presented a novel decision-making paradigm focused on an IoT method for identifying and tracking type 2 diabetes patients. Wireless BAN was used to track changes in the user's body symptoms, and a smartphone phone interface was used to record social interactions. Since it was necessary to enhance the decision support schedules for the accurate predictions of type 2 diabetes issues, the hybrid approach focused on type 2 neutrosophic with the VIKOR process were proposed in this analysis. The performance of this model was satisfactory and the accuracy could be improved by using advanced approach [6]. Abdur RMF et al. proposed a knowledge discovery-based approach that enabled a context-aware system to change its behaviour in real time by analyzing large volumes of data produced in ambient assisted living frameworks and stored in the cloud databases. The proposed model allowed big data research within a cloud setting. It first analyzed the dynamics and patterns in a particular patient's records, along with the associated odds, and then used the information to learn proper irregular conditions. The results of this learning approach were then used in context-aware decision-making scheme for the patient. This model can be improved with more context domains [7]. Deeba K and RA. K. Saravanaguru proposed a model for monitoring signs and health conditions of elderly people. The data from the system was observed by the caregivers for identifying the daily activities through IoT. A fuzzy logic controller was designed from the initial stage of data collection, data processing, filtering and accumulating it into contextual data and reasoning for identifying the elder people's health conditions [8]. Jalil N-K et al. proposed a novel hybrid technique for heart diseases diagnosis using optimization method in feature selections. This analysis mainly focused on the features selection enhancement and reducing the features count. In this analysis, imperialist competitive algorithm with meta-heuristic technique was proposed to choose essential features of the heart diseases and the K-nearest neighbour technique was utilized for the classification. This model could enhance the features selection technique for missed and incomplete data [9].

Daniel A D et al. designed a context-aware system to assist health care providers in home-based caring environments. A reliable NFC authentication scheme was used, which creates a secure channel by encoding sensitive contextual data during data transmissions. Using a context-aware gateway node, this system performs authentications and authorization for accessing a specific patient's data. The proposed solution aimed to improve health care data access and safe data delivery while protecting users' privacy. This research provided a foundation for physicians to develop different smart treatment alternatives, as well as for home-based care [10]. Deeba K and RA. K. Saravanaguru proposed a Smart Home Caregivers System (SHCS) capable of collecting real-time patient's heart rate, oxygen leakage in, abnormal and normal patient's condition observed through MQ6 sensor. The data sensed was transmitted to the base station, where it was controlled by caregivers via PC or mobile device. This method was carried out by either wired or remote users using REST web services [11].

Based on Fuzzy Logic, Byung-K L et al. presented context-aware health care model for disease reasoning. It was made up of two modules: Fuzzy-based disease reasoning model (FDRM) and the Fuzzy-based context aware model (FCAM). The FCAM calculated the correlations coefficients and supports among the conditional attributes and the decision attributes and produced fuzzy rule based solely on the conditional attributes with the highest correlation coefficients and supports. Based on the results performed with a SIPINA mining method, the average accuracy of Fuzzy Rules dependent on correlations coefficients and supports (FRCS) and enhanced C4.5 was 0.84 and 0.81, respectively. That was, as correlated to the enhanced C4.5, the FRCS reduced the rules produced while improving accuracy of rules [12].

3. Proposed Methodology

In this work, a context aware health care method based on the application of IoT was proposed. Smart medical devices were utilized to collect and retain patient data, which was stored in a database. The database contained context-aware data, such as the names, addresses, and medical histories of the patients. A rule-based machine learning technique, a modified RIPPER algorithm, was utilized to analyze and classify the data. The rules for analyzing data for the prediction of heart disease were developed using this algorithm. Based on the classification results, the prediction of the heart disease was performed (Fig. 2).

The smart medical wearable devices based on body sensor network is used for the computation of patient's physiological medical data (i.e., Heart rate, temperature, etc.). The data from these devices are stored in the cloud platform for data storage. The stored data can be further managed or used by the user or the medical centres for analyzing the patient's health conditions. By using specific application, patient can be monitored from remote places via internet through smartphones or PCs. The block diagram of the model was shown in figure.2. The proposed algorithm is discussed as follows.

The IoMT devices and wearable devices are considered as the IoT devices. They are equipped to accumulate the patient's data from remote areas. These data are collected as patient's information that are accumulated using IoT devices connected or equipped with the human body.

3.1. RIPPER ALGORITHM

The rule-based machine learning can be detailed as the basic concept description. The RIPPER algorithm is one among them most widely used. Comparing to various algorithms this technique has more benefits, that it could be comprehended with ease, with generated rules in the form of If-Then format, implying that the model is entirely interpretable. The RIPPER algorithm is a rule-based classification algorithm that produces a rule-based classifier model, which is a collection of IF-THEN rules derived straight from the training data set, thus the name "direct process." It can be utilized for multi-class and binary classifications. The RIPPER algorithm's core framework was split into two types: optimization rules and generation rules. The generation type is the two-layer loops in which the outer loop produces the rule and applies it to the rule base next to pruning, while the inner loop includes one antecedent to the rule at once. The optimization type creates alternate rules based on the rule base's rules, and the minimum description length (MDL) criteria was utilized to pick the right rule and attach with rule base [13]. This algorithm goes through four stages:

Growth

During this process, a rule was created by greedily applying features to the rule before it passes the stopping criterion.

Pruning

Throughout this process, every rule was pruned and rendered shorter by eliminating repetition and reducing the duration of previous rules, allowing the rule to improve.

Optimization

The initial prune and growth process creates rules from an empty ruleset. The optimization phase makes use of the rules created during the initial pruning and growth stages and attempts to create new rules from the rule set. The rules can be additionally optimized with,

Adding features to the initial rule using the greedy approach (i.e., depth initial search).

Following the growth and pruning process, a new rule set is created.

Selection: At the selection process, the best rules were held and the rest of the rules are removed from the system. The specifications of this algorithm are as follows: D dataset is used as input (Eqs. 1 and 2).

Step 1

Split the dataset D into individual growth sets Gro and prune set Pru

Step 2

The growth set Gro was utilized at this point as dataset. The growth rule starts with no rules, and every time an appropriate combination of potential features and thresholds are chosen as the antecedents would be included to the rules. The information gain was utilized as the evaluation criterion

$$IGN = \text{cover}(\log_2 rt' - \log_2 rt)$$

1

Cover is the number of positive instances that were covered since adding the antecedents to the rule, *rt'* was the proportion of positive instances in the data covered using the rule since adding the antecedents to a rule, and *rt* is not. The iteration of including antecedents would continue until the *Gro* was empty.

Step 3

The pruning process utilize the pruning set *Pru* to measure the rule's generalization capacity. Begin with past thing added, and eliminate an antecedent in the rule. When pruning, the metric was

$$T = \frac{p - n}{p + n}$$

2

Table.1. Description of Data used in Ruleset

Type	Range	Description
Chest Pain	1 to 4	Typical angina
		Atypical angina
		Non angina
		Asymptomatic
Cholesterol	< 197	Lower
	188–250	Medium
	217–307	Higher
	> 281	Very higher
BP	< 134	Lower
	124–153	Medium
	142–172	Higher
	> 154	Very higher
Blood Sugar	< 120	No
	>=120	Yes
ECG	< 0.4	Normal
	0.4–1.8	Abnormal
	> 1.8	Hypertrophy
Thallium	3	Normal
	6	Fixed Defect
	7	Reversible Defect
Age	< 35	Younger
	35–45	Middle
	40–58	Older
	> 58	Very older
Gender	1	Male
	0	Female
Smoking (in years)	<=10	Lower
	> 10	Higher
Drinking	0	No
	1	Yes

Type	Range	Description
Family history (diabetes, hypertension, ...)	< 1 >= 1	No Yes
Medical records (diabetes, hypertension...)	< 1 >= 1	No Yes

p was positive instances covered by the rule in Pru , n was negative instances covered by the rule in Pru , the point of the calculation was to increase the precision of the pruned set.

Step 4

After pruning the rule, it was tried to be included to a rule base. The inclusion would fail if the number of instances covered by the rule was too limited or the precision was too poor. If the rule was effectively included, the instances covered will be removed from the D [14].

3.2. MRIPPER ALGORITHM

The RIPPER algorithm for rule induction was implemented as a replacement for the Incremental Reduced Error Pruning (IREP) algorithm. About the fact that the fundamental ideals remain similar, Modified-RIPPER strengthens IREP in certain details and was also capable of dealing with multiclass issues. A single MRIPPER rule was made up of a consequent and an antecedent part. The antecedent part was a predicate (selector) conjunction, and the consequent part was a class assignment. MRIPPER learns those rules greedily, using a divide-and-conquer approach. The training data are classified by class terms in increasing order based on the respective class frequencies prior to the learning process. The rules for the initial $m-1$ classes are then learned, beginning with the smallest. When the rule was established, the instances concealed by that rule are excluded from the training data, and this process was replicated till no instances from the target classes remain. After that, the algorithm moves on to the next class. Finally, as MRIPPER discovers that there are no more rules to learn, a default rule (with an empty antecedent) was applied for the last class. Single-class rules are learned before either all positive instances were concealed or the last rule applied was "too difficult." The last feature was applied in terms of overall description length: the stopping criteria was met if R's description length was quite longer than the shortest description length found so far as represented in the algorithm [15].

3.3. PROPOSED ALGORITHM

procedures BUILDSET (P, N)

P = *positive samples*

N = *negative samples*

Rule Set = \emptyset

DL = *Description length (Rule Set, P, N)*

```

while P {

//Grow and prune a new rule

split (P, N) into (Gro P, Gro N) and (Pru P, Pru N)

Rule = Gro Rule (Gro P, Gro N)

Rule = Pru Rule (Rule, Pru P, Pru N)

add Rule to Rule Set

if Description Length (Rule Set, P, N) > DL + 11 then

//Prune the whole ruleset and exit

for every rule R in Rule Set (considered in reverse order)

if Description Length (Rule Set {R}, P, N) < DL then

delete R from Rule Set

DL = Description Length (Rule Set, P, N)

end if

end for

return (Rule Set)

end if

DL = Description Length (Rule Set, P, N)

delete from P and N all instances covered by Rule

end while

end BUILDRULESET

procedure OPTIMIZERULESET (Rule Set, P, N)

for every rule R in Rule Set

delete R from Rule Set

U Positive = instances in P not covered by Rule Set

U Negative = instances in N not covered by Rule Set

```

split (UP, UN) into (Gro P, Gro N) and (Pru P, Pru N)

Rep Rule = Gro Rule (Gro P, Gro N)

Rep Rule = Pru Rule (Rep Rule, Pru P, Pru N)

Rev Rule = Gro Rule (Gro P, Gro N, R)

Rev Rule = Pru Rule (Rev Rule, Pru P, Pru N)

select best of Rev Rule and Rep Rule and add to Rule Set

end for

end OPTIMIZERULESET

procedure RIPPER (P, N, k)

Rule Set = BUILDRULESET (P, N)

repeat k times

Rule Set = OPTIMIZERULESET (Rule Set, P, N)

return (Rule Set)

end RIPPER

3.4. RULESET

Rule 1: if (Thallium was less) and (Chest pain was typical angina) hence (No heart disease)

Rule 2: if (Thallium was less) and (Chest pain was atypical angina) hence (No heart disease)

Rule 3: if (Thallium was less) and (Chest pain was non angina) hence (No Heart disease)

Rule 4: if (Thallium was less) and (Chest pain was asymptomatic) and (Vessel was less) hence (No Heart disease)

Rule 5: if (Thallium was less) and (Chest pain was asymptomatic) and (Vessel was higher) hence (Presence of Heart disease_2)

Rule 6: if (Thallium was higher) and (vessel was less) and (No angina) hence (No Heart disease)

Rule 7: if (Thallium was higher) and (vessel was less) and (has angina) hence (Presence of Heart disease)

Rule 8: if (Thallium was higher) and (vessel was less) hence (Presence of Heart disease_2)

Rule 9: if (family history was true) hence (Presence of Heart disease_1)

Rule 10: if (medical record was true) hence (Presence of Heart disease_1)

Rule 11: if (family history was true) and (medical record was true) hence (Presence of Heart disease_1)

The ruleset for the algorithm discussed is framed based on the medical condition of the patient regarding the prediction of heart disease prediction. The ruleset is framed from the conditions related to the causes of heart disease as shown in table.1. Each attribute represented in the table has the range based on the condition of description. According to the ranges, the disease condition will be predicted with the ruleset framed for context aware system. The context database collects the context aware data from the input which is medical records, personal information, and physiological data of the patient. Based on the context analyzer and context reasoning the ruleset is framed and analyzed. The data from the context database is processed by the proposed algorithm for the classification. For experimental analysis, the Cleveland dataset is implemented in this research for evaluation.

4. Performance Analysis

The proposed model's performance analysis was experimented in the MATLAB Simulink tool. The experiment was carried out on a 64-bit CPU, i5 processor operating on Windows 10, with 8 GB RAM, using the MATLAB & Simulink tool R2017a. The data classification is significant in this analysis. The proposed classifier classifies the data for the prediction of heart disease, which the result will be in the form of absence or presence of disease. The results are carried out using the dataset, and various classification parameters like accuracy, recall, precision, and F-measure. For classification, the benchmark dataset is classified using rule-based machine learning classifier called modified RIPPER algorithm. In this analysis the heart disease prediction is the main work concentrated and the prediction model can be used to perform prediction for any serious disease by using the different dataset in the process.

4.1. DATASET DESCRIPTION

For the heart disease prediction and classification, Cleveland data set from UCI repository was utilized in this research, available in public database. Each data set has its very own instances and attributes, in that Cleveland dataset is used for training which has 76 attributes and 303 records. But, only 13 attributes in dataset were utilized for this analysis & experiment as represented in Table 2 [16].

Table.2. Descriptions of Cleveland Dataset

Medical Term	Description
Age	Age (years)
Gender	1 = male; 0 = female
Chest pain	Types of chest pains: 1- typical angina, 2- atypical angina, 3- Non-anginal pain, 4- Asymptomatic
Bps	Resting blood pressure (mm HG)
Chol	Serum cholesterol (mg/dl)
Fastbs	Fasting blood sugar > 120mg/dl: 0-false; 1-true
Continuous Max heartrate measured	Exercises induced angina: 0-No; 1-Yes
Thalac	Max heartrate obtained
ST	Depressions induced by exercises related to rest
Slopes	The slopes of the peak exercise segments: 1-up sloping, 2-flat, 3-down sloping
Cal	Total major vessels coloured by fluoroscopy which ranged among 0 and 3
Thall	3-Normal, 6-Fixed defects, 7-Reversible defects
Classes	Diagnosis classes: 0-Healthy, 1- Presence of heart disease.

Table.3. Sample of Dataset

Age	Sex	Cp	Trestbps	Chol	Fbs	Induced angina	Thalach	ST	Slope	Ca	Thal	Class
55	0	3	115	322	0	0	160	1.6	2	0	7	0
74	1	2	124	261	0	0	141	0.3	1	0	7	1

The data sample from the Cleveland dataset is shown in table 3. It represents the data of two patients collected and in the form of numerical value. Here the medical term data are represented numerically according to the description presented in table 2. Performance metrics such as accuracy, sensitivity, specificity, precision, and f-measure are used to evaluate the proposed model's performance. These metrics are calculated using true positive (Tpos), true negative (Tneg), false positive (Fpos), and false negative (Fneg) measurements.

In general, the accuracy of data is determined by how it was obtained. It was calculated by comparing various metrics from the same or different source (Eq. 3).

$$\text{Accuracy} = \frac{Tpos + Tneg}{Tpos + Fpos + Tneg + Fneg} \% \quad (3)$$

Precision, also known as positive predictive values, was the probability that a person with a positive screening test truly had the disease. The precision may be determined, as stated by the Eq. 4.

$$\text{Precision} = \frac{T_{pos}}{T_{pos} + F_{pos}} \quad (4)$$

Recall or Sensitivity measures the ability to detect a patient at risk for heart disease and is stated as an Eq. 5.

$$\text{Recall} = \frac{T_{pos}}{T_{pos} + F_{neg}} \quad (5)$$

Specificity is measured by dividing the total number of negatives by the number of genuine negatives, as shown. The highest specificity is represented by 1.0, while the lowest is marked by 0.0. (Eq. 6)

$$\text{Specificity} = \frac{T_{neg}}{T_{neg} + F_{pos}} \quad (6)$$

The F-measure, which is defined as the weighted harmonic mean of test precision and recall, assesses test accuracy. The accuracy does not take into account how the data is disseminated. As a result, the f-measure is utilized to accurately manage the distribution problem (Eq. 7).

$$\text{F-Score} = \frac{2T_{pos}}{2T_{pos} + F_{pos} + T_{neg}} \quad (7)$$

T_{pos}-True Positive is proper prediction on healthy classes;

T_{neg}-True Negative is proper prediction on abnormal classes;

F_{pos}-False Positive is improper prediction on healthy classes;

F_{neg}-False Negative is improper prediction on abnormal classes.

From the dataset, the selected data are analyzed and classified by the proposed MRIPPER algorithm. Five healthy samples and five abnormal (Heart disease) samples are used for the experiment in this research. The subject 1–5 are classified as healthy class and 6–10 are classified as abnormal class from the dataset. The performance analyses of these ten subjects are computed and tabulated in table 4. Figures 3 and 4 represents the graphical representation of performance analysis made on normal and abnormal subjects as shown in table 4. Accuracy, precision, recall, specificity, and f-score are the parameters used in this research for the evaluation of performance.

Table 4. Comparison of Performance Evaluation on Normal and Abnormal Subjects using MRIPPER algorithm.

Class	Accuracy	Precision	Recall	Specificity	F-Score
Healthy Class	96.48	94.81	97.21	89.64	95.50
	97.26	95.20	98.43	90.26	96.85
	96.40	94.73	97.10	89.12	95.08
	97.91	95.08	98.43	91.35	96.32
	98.85	96.93	99.20	92.48	97.05
Abnormal Class	98.97	94.99	99.62	91.95	97.37
	98.51	96.64	99.45	90.16	97.01
	97.32	95.82	98.35	89.78	96.95
	98.04	96.34	98.92	91.65	97.46
	96.85	94.75	97.45	89.51	95.68

Table.5. Comparison of Performance Analysis

Algorithm	Accuracy	Precision	Recall	Specificity	F-Score
J48	94.08	93.45	95.82	90.24	91.18
Random Forest	95.56	93.83	94.20	91.70	93.55
CART	95.80	94.06	96.48	90.96	96.13
OneR	96.48	93.21	96.54	92.11	95.24
JRip	97.66	95.01	97.80	93.54	96.15
MRIPPER	98.89	96.76	99.05	94.35	97.60

Table 5 represents the comparison of the proposed algorithm's performance analysis with other existing algorithms. The proposed MRIPPER algorithm is compared with random forest, J48, CART, JRip, and OneR algorithms. The proposed approach obtained 98.89 percent accuracy, which was 1.2–4.8% higher than the other algorithms. Precision of the MRIPPER algorithm is 96.76%, which is 1.7–3.5% higher, recall or sensitivity is 99.05%, which is 1.2–4.8% higher, specificity is 94.35%, which is 0.8–4% higher, and f-score is 97.60%, which is 1.4–6.4% higher than the other algorithms. Overall, the proposed approach has attained 97.93 percent in predicting abnormal class and 97.38 percent accuracy for normal class subjects (Figs. 5 and 6).

5. Conclusion

The use of context awareness in medical field is embedded with other domains like IoT, Cloud computing, etc. With the combination of integrating with these technologies the developed application on context awareness system has many advantages over other methods. Different applications like health monitoring, analyzing on diseases, and assisting on medications can be done on remotely with the combinations of these

technologies. In this work, a context aware health care method based on the application of IoT was proposed. Smart medical devices were utilized to collect and retain patient data, which was stored in a database. The database contained context-aware data, such as the names, addresses, and medical histories of the patients. A rule-based machine learning technique, a modified RIPPER algorithm, was utilized to analyze and classify the data. The rules for analyzing data for the prediction of heart disease were developed using this algorithm. Models like random forest, J48, CART, JRip, and OneR algorithms are compared for the validation of the proposed model. The proposed model obtains 98.89 percent accuracy, 96.76 percent precision, 99.05 percent sensitivity, 94.35 percent specificity, and 97.60 percent f-score. Predictions for subjects in the normal and abnormal classes were both accurate with 97.38 for normal and 97.93 for abnormal.

Declarations

- Funding: Yes
- Conflicts of interest/Competing interests: The authors declare no conflict of interest, financial or otherwise.
- Availability of data and material: The authors confirm that the data supporting the findings of this research are available within the article.
- Code availability: Custom code
- Authors' Contributions: There are seven authors in this article, and all are contributed equally.
- Human And Animal Rights: No animals/humans were used for studies that are basis of this research.
- Ethics approval: Not applicable.
- Consent to participate (include appropriate statements): Not applicable.
- Consent for publication (include appropriate statements): Not applicable.

ACKNOWLEDGMENT

- This work was financially supported by Industrial Innovation & Robotics Center, University of Tabuk, Tabuk City, Kingdom of Saudi Arabia. Research Project Number: 0186-1441-S.
- An earlier version of this work has been presented as a preprint by Research Square [17].

References

1. Yousef A (2020) A Context Aware Framework for IoT Based Health Care Monitoring Systems. *Int J Adv Stud Comput Sci Eng* 9(7):1–9
2. Rakshitha P, Ashok I (2017) A Survey on Context Awareness Security in Health care. *Int J Adv Res Comput Sci* 8(3):219–222
3. Pooja SG, Chaware SM (2018) Context Aware Computing System: A survey. Proceedings of the Second International conferences on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), pp.605–608
4. Luis CG, Cristiano AC, Rodrigo RR (2019) Context awareness in health care: a systematic literature review. *Univ Access Inf Soc*. <https://doi.org/10.1007/s10209-019-00664-z>

5. Reddy RV, Murali D, Rajeshwar J (2019) Context-Aware Middleware Architectures for IoT-Based Smart Health Care Application. H. S. Saini (eds.), Innov Comput Sci Eng, Vol.74, pp.557–567
6. Mohamed A-B, Gunasekaran M, Abdullaah G, Victor C (2019) A Novel Intelligent Medical Decision Support Model based on Soft Computing and IoT. IEEE IoT J 7(5):4160–4170
7. Abdur RMF, Ibrahim K, Ayman I, Zahir T, BDCaM Big Data for Context-aware Monitoring- A Personalized Knowledges Discovery Frameworks for Assisted Health Care.IEEE T Cloud Comput, 5(4):628–641
8. Deeba K, Saravanaguru RAK (2020) Context-Aware Elderly People Monitoring based on IoT. J Xi'an Univ Arch Technol 7(3):5797–5804
9. Jalil N-K et al (2019) New hybrid method for heart diseases diagnosis utilizing optimization algorithms in feature selections. Health Technol. <https://doi.org/10.1007/s12553-019-00396-3>
10. Daniel AD et al (2019) Privacy Enhanced Health Care Information Sharing Systems for Home-Based Care Environment. Healthc Inf Res 25(2):106–114
11. Deeba K, Saravanaguru RAK (2018) Context-aware Health Care System based on IoT- Smart Home Caregiver Systems (SHCS). Proceedings of the Second International Conference on Intelligent Computing and Control System, pp.1204–1208
12. Byung -KL, Eun -HJ, Sang -SL (2016) Context-Awareness Health Care for Diseases Reasoning Based on Fuzzy Logic. J Electr Eng Technol 11(1):247–256
13. Sonika T, Meenakshi (2020) Detection of Malicious URL in Big Data using Ripper Algorithm. Int J Sci Eng Res 11(4):61–66
14. Aruna G, Varsha ST, Ipsita S, Sanjay KS (2016) Distributed multi-class rule-based classifications using RIPPER. IEEE International Conference on Computer and Information Technology, pp.303–309
15. Rajeswari M, Devi T (2015) Design of modified ripper algorithm to predict customer churn. Int J Eng Technol 4(2):408–413
16. <http://archive.ics.uci.edu/ml/datasets.php>
17. https://assets.researchsquare.com/files/rs-919935/v1_covered.pdf?c=1632855159

Figures

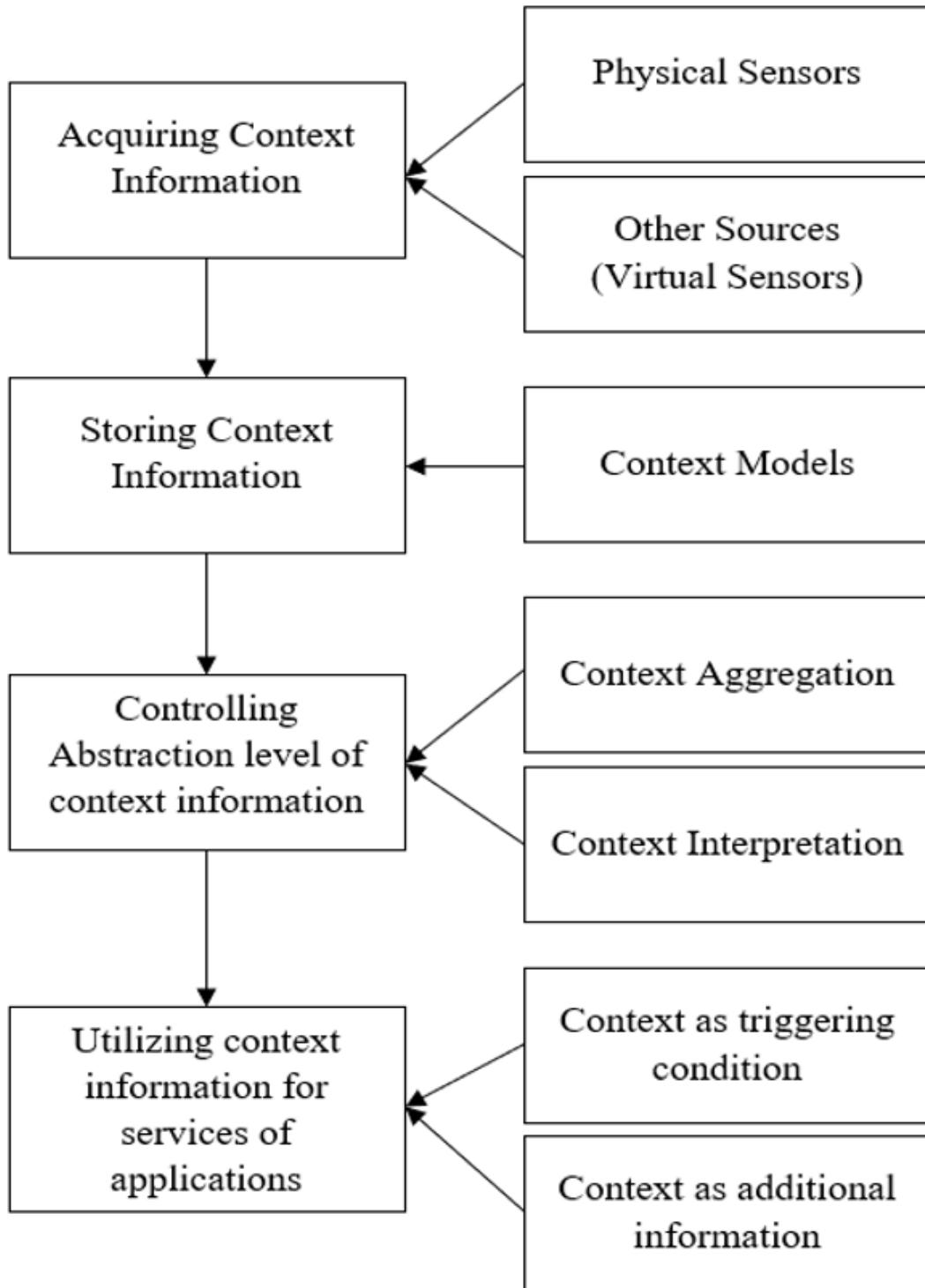


Figure 1

General Process of Context Aware System

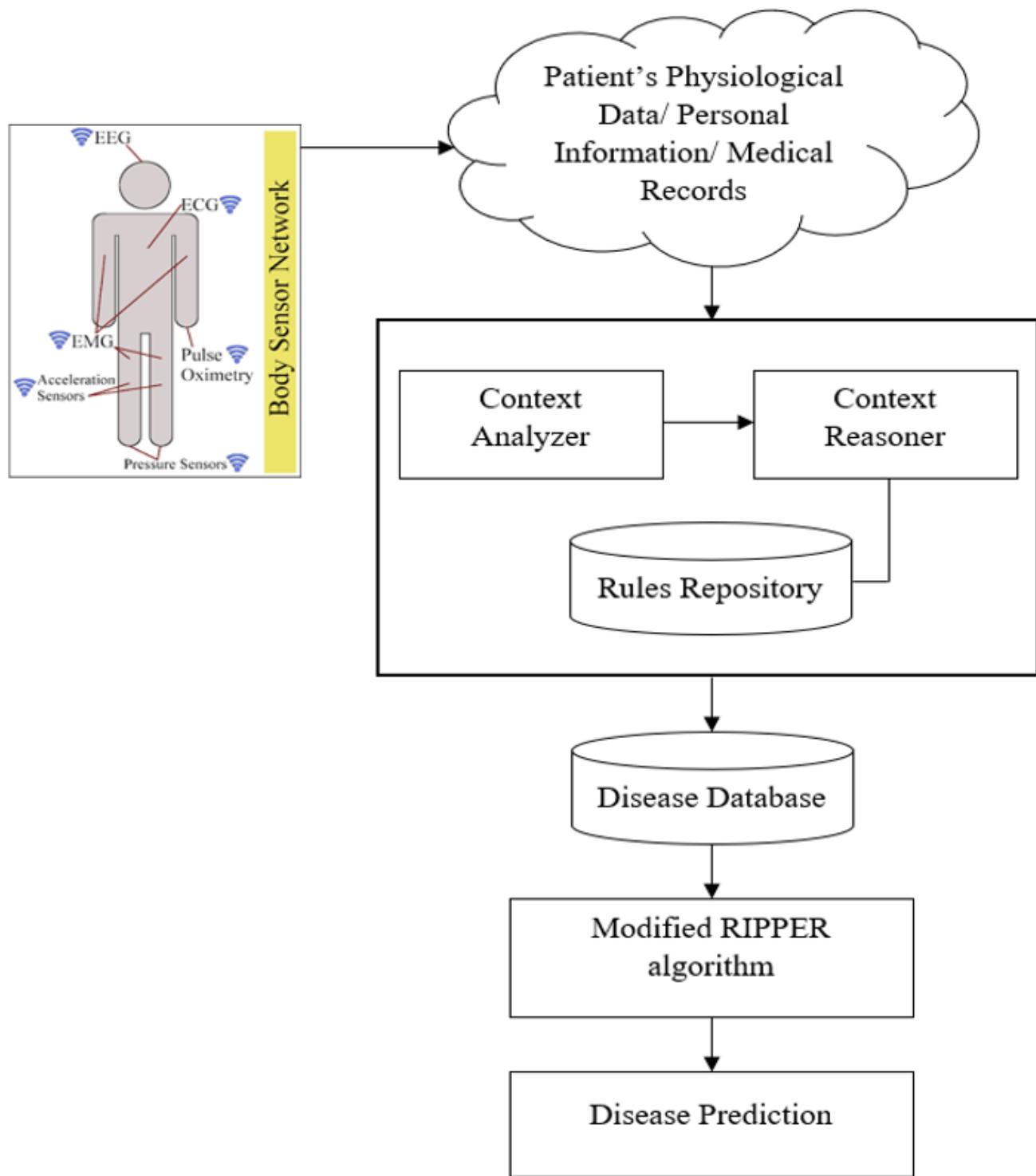


Figure 2

Block Diagram of the Proposed Model

NORMAL SUBJECTS

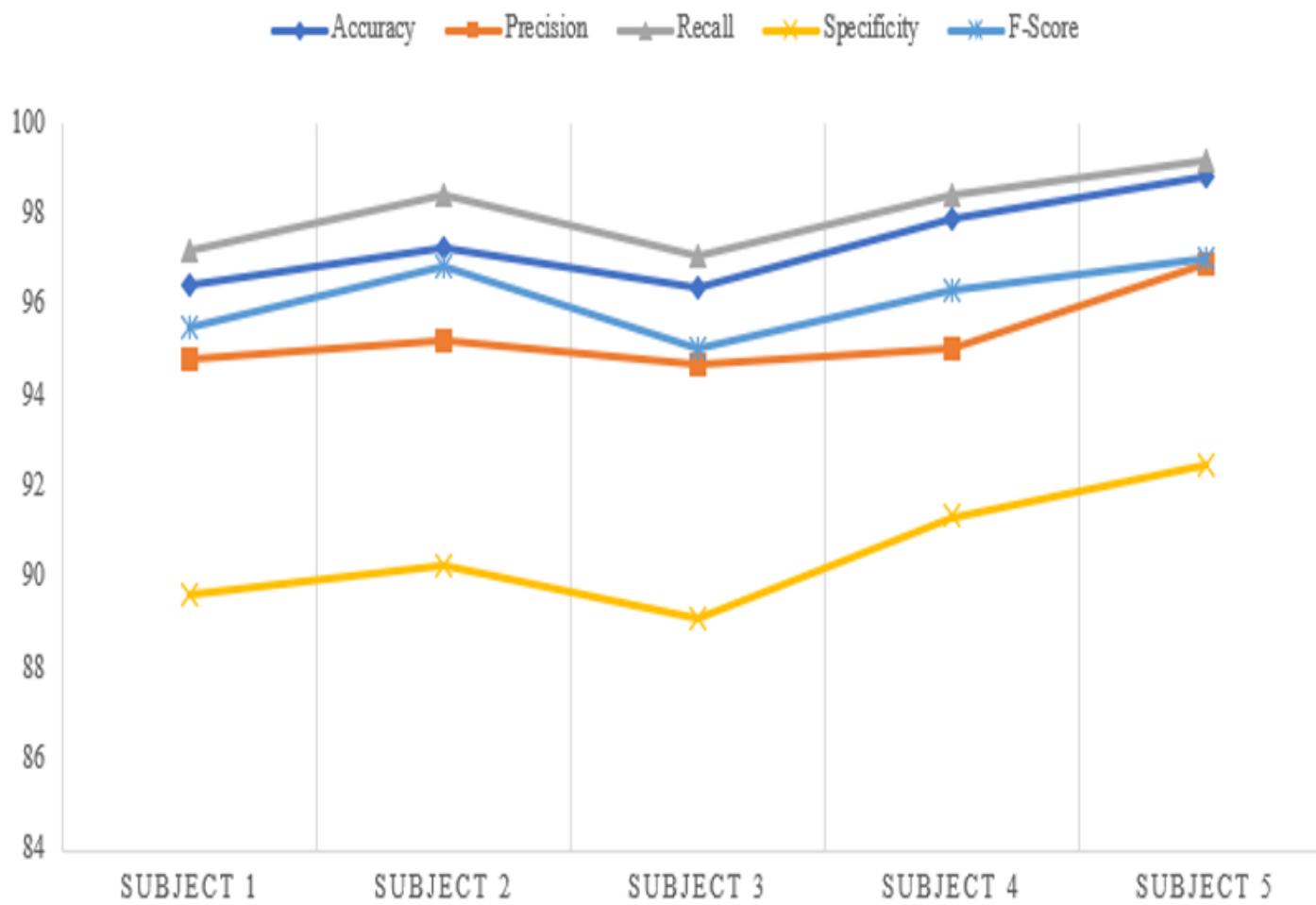


Figure 3

Performance Analysis on Normal Subjects

ABNORMAL SUBJECTS

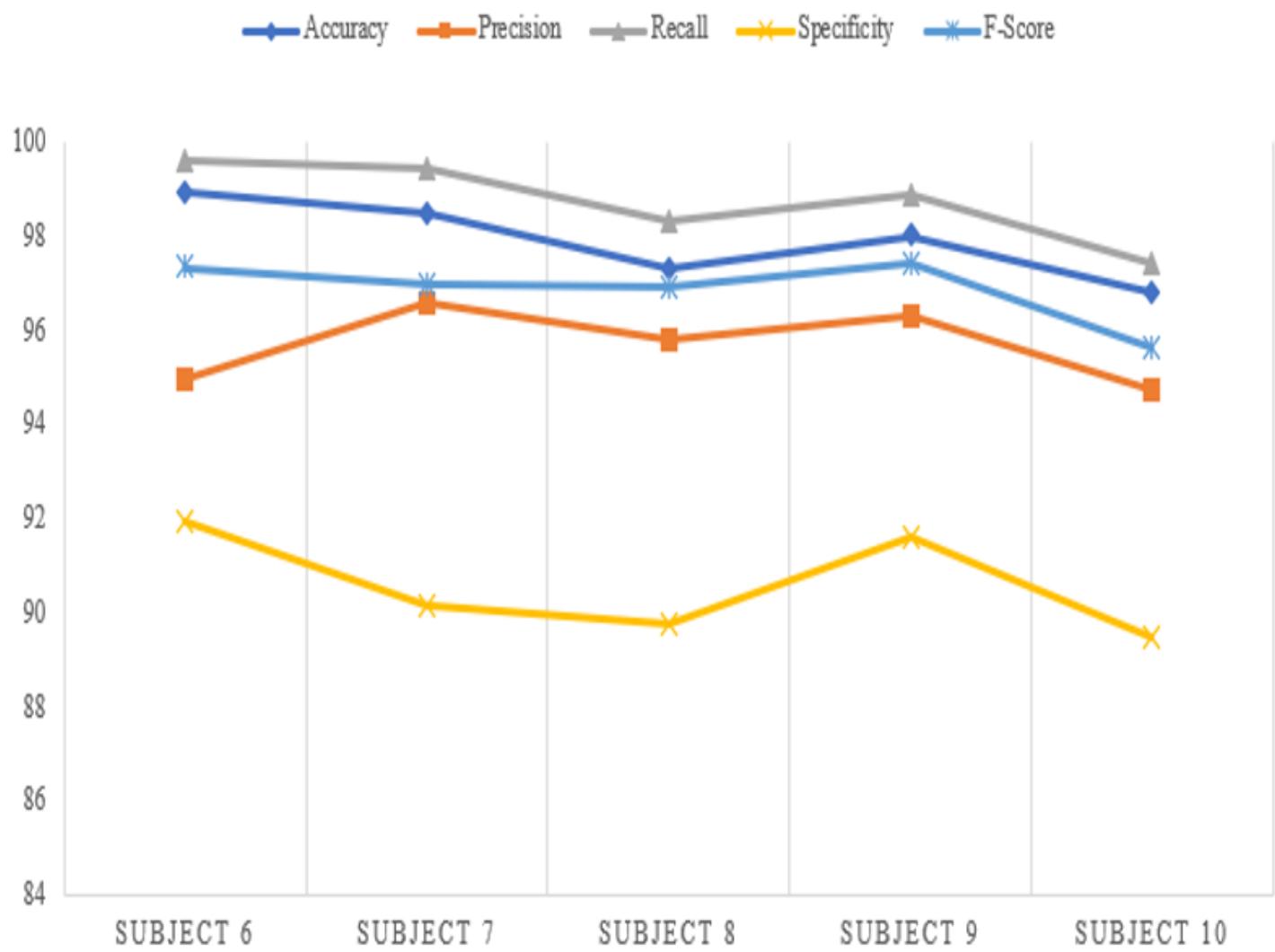


Figure 4

Performance Analysis on Abnormal Subjects

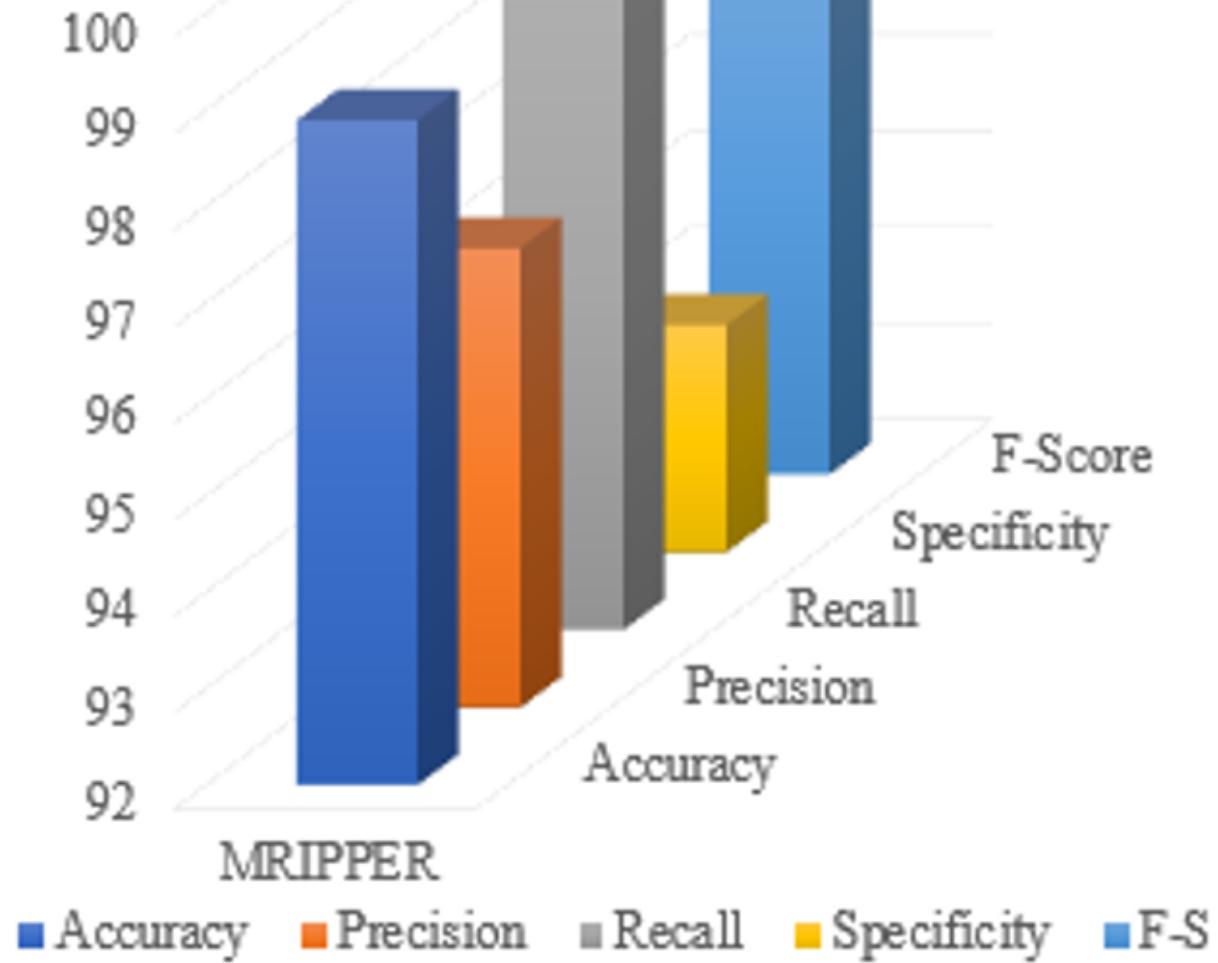


Figure 5

MRIPPER Performance

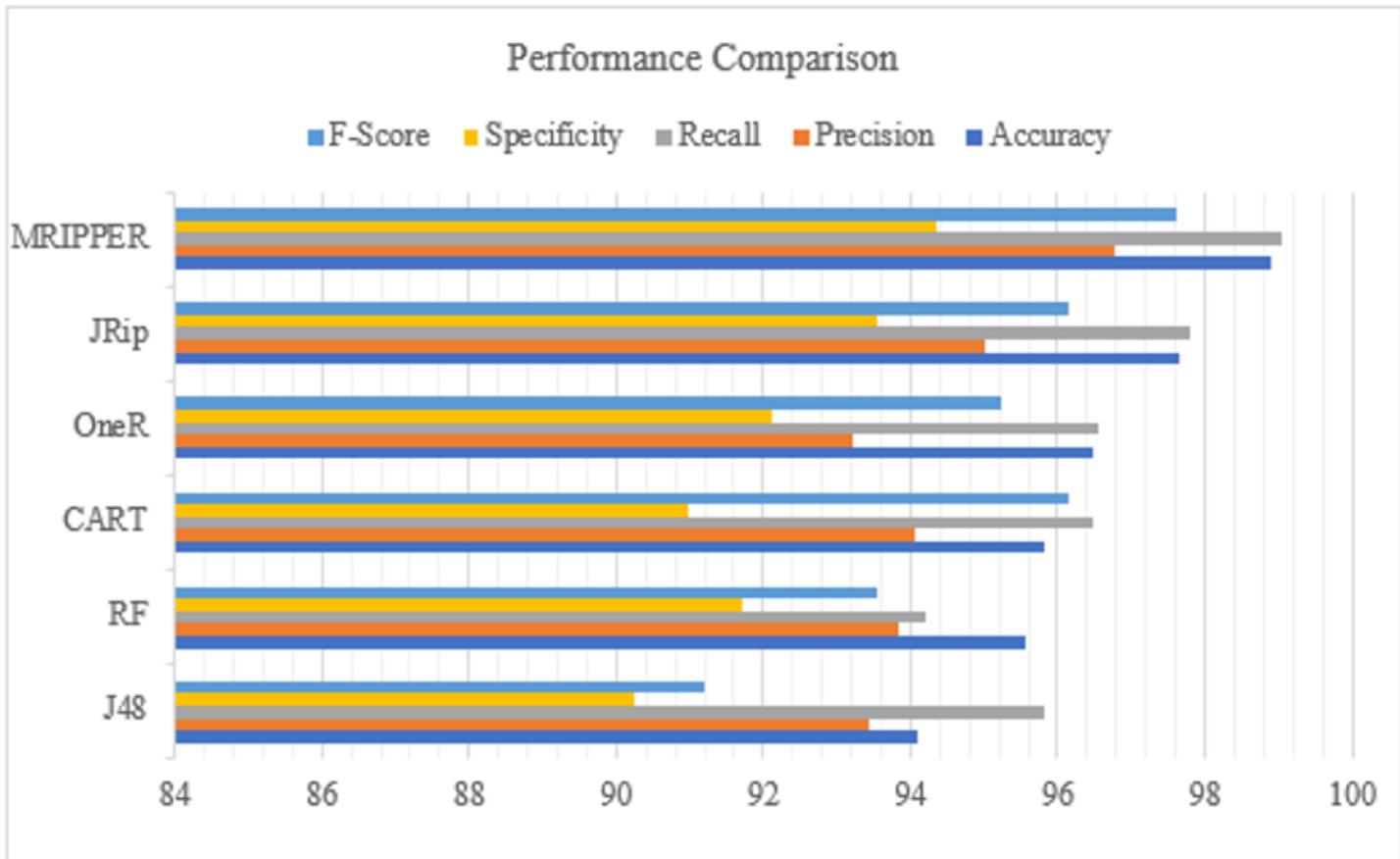


Figure 6

Comparison of Performance with Existing Algorithms