



A Descriptive Study of Predictive Models of MERS-CoV Outbreak

Anjali Saini
Department of computer science
GNDU, Regional Campus
Jalandhar, India

Dr. Pankaj Deep Kaur
Department of computer science
GNDU, Regional Campus
Jalandhar, India

Abstract: Currently, Medical care has not completely appreciated the possible advantages to be acquired from Big Data Analytics. The evolving areas of efficient utilization of Analytics are predictive analytics and traceability. The Big Data Analytics abilities and possible benefits are provided in healthcare to formulate far better data-driven analytics techniques for MERS-CoV disease. This paper is basically giving emphasis on the MERS-CoV. It is an airborne illness which develops simply and has large death rate. It could be the acute respiratory syndrome brought on by betacoronavirus and belongs to the coronaviruses household, which is responsible for producing gentle to reasonable colds. The initial situation of the illness was reported from Saudi Arabia. Recently, the outbreak of MERS-CoV infections triggered global attention to Saudi Arabia. Relating to the records, males are far more susceptible than girl, especially following the age of 40. Due to the understanding and early diagnosis the incidence is slipping gradually. There is no unique therapy for the MERS-CoV till now. As a representative case study the Data Mining techniques are proposed that can be used in order to better realize the security and the chance of recovery from MERS-CoV infections. In addition unsupervised filtering is used for better performance and accuracy of predictive Models.

Keywords: Big Data; Prognosis; MERS-CoV; Data Mining; Big Data Analytic; Decision tree; J48; Naïve Bayes; Classification.

I. INTRODUCTION

Big Data refers to large amounts of knowledge, made at very high rates and a huge number of forms. It depends on a set of highly per formant technologies combined together to take advantage of these unprecedented amounts of varied and fast data. Big Data is a new area that encompasses numerous disciplines and affects a wide variety of sectors of our society. Its rapid increase in new years could be linked to many scientific advances. The raising option of sensors produced data technology and series easier and cheaper [2]. Developments in telecommunications systems and services facilitated the enormous exchange of data among client devices, data stores and clouds. The quickly lowering of data storage and control charges offers increase to quickly development in raising computational power. As a result, novel programs are widely found that amount across varied fields as never before. The main features of Big Data are categorized as follows: Volume, Velocity, Variety, Value, and Veracity.

Big data and analytics are topics that are capturing attention from worldwide now days. Gathering and holding large data produces little value; it is only data infrastructure as of this point. It must certainly be analyzed and the outcomes utilized by choice makers and organizational processes to be able to make value. Many analytic practices, such as regression analysis, simulation, and device understanding, have already been readily available for several years. Even the worth in considering unstructured data such as email and papers has been properly understood. What is new is the coming together of developments in pc technology and pc software, new sources of data (e.g., social media), and business opportunity. This confluence has created the current curiosity and possibilities in major data analytics. It's even spawning a new section of training and examine named "data science" that encompasses the practices, methods, systems, and procedures in making sense out of major data.

Big Data Analytics is targeted at creating feeling of data by applying effective and scalable methods on Big Data for their evaluation, learning, acting, visualization and understanding. This includes the design of effective and successful methods and techniques to incorporate the data and uncover the concealed prices from data. It also incorporates methodologies and methods for automated or mixed-initiative understanding finding and learning, data transformation and acting, predictions and explanations of the data. Breakthroughs in this region include new methods, methodologies, techniques and purposes for understanding finding, understanding and purposes based on the Large Data. New research paradigms are expected in new parts such as for instance individual computation, crowd sourcing, and feeling evaluation along with data visualization technologies. Big data analytics is a fast-growing and influential practice [5].

Data mining is some functions meant to transform raw data into new and helpful knowledge. Normal guidelines have already been identified for applying a data mining process. When applying a data mining project, it is important to follow a regular and proven system in order to avoid popular traps and lower risks

There are lots of strategies open to mine large data pieces for important information. From visualization to the absolute most sophisticated device learning algorithms, you will find various instruments to remove understanding from data [3]. Common device learning algorithms could be placed on typical issues in energy systems. All algorithms use a data dining table as an insight where in actuality the rows are named the items (a snapshot of the indication grid state recognized by way of a timestamp for example) and the articles are named the attributes (e.g., the voltages levels, shows, and currents in lines of the indication grid state). Equipment learning algorithms are generally divided into two main types of algorithms: monitored and unsupervised [14]. Unsupervised algorithms are useful when you don't have a

specific productivity variable to spell out or predict. Generally, this category of algorithm is employed for:

1. Link analysis problems to discover correlations or associations between attributes. Link analysis would be used to discover a group of measurements that are strongly correlated, such as voltages of the nodes located in an area.
2. Clustering problems to detect clusters of objects that have similar behaviour, such as states of the power grid that are similar.

There are number of tasks related to mining that are Classification, Association rules, Clustering and Anomaly detection. But the main focus of paper is on Classification using J48 Decision tree. Decision Tree involves a set of practices targeted at generating detailed tree structures that classify files by selecting them centred on attribute values. Each node in a decision tree represents an attribute in a record to be labelled, while each part presents a benefit that the attribute can take. Base Remaining for an aesthetic example [15]. Choice woods where the goal variable can accept constant prices are named regression trees. Let D be the set of training records that reach a node. The general procedure to build the tree is as follows:

- If D contains records that belong to the same class, then this is a leaf node.
- If D contains records that belong to more than one class, use an attribute test to split the data into smaller subsets. Apply the procedure recursively on the obtained subset.

J48 Decision Tree is one of the famous data mining techniques. It is definitely an implementation by the WEKA task staff of the well-known tree induction algorithm C 4.5. It uses a greedy iterative strategy in developing the decision tree. The algorithm surfaces the dataset based on the most effective educational attribute. At each iteration, the feature with maximum get percentage is selected whilst the breaking attribute [1]. Decision tree classification types have several advantages. They are easy to understand and are recognized to have comparable precision to other classification models. In the J48 decision tree healing product, the attribute healthcare workers appear as the first dividing attribute. This indicates the significance of this information. The model may be viewed the following: if the patient is a healthcare worker, the product anticipates recovery. But, if the in-patient is not really healthcare personnel then your product examines whether he has any pre-existing disease. If the in-patient suffers from other disorders, the product anticipates demise, usually healing is predicted. According to this model, healthcare workers are more prone to survive MERS-CoV infections .There are two models Stability and recovery models as follows:

Figure 1.3 J48 Decision Tree Recovery Model

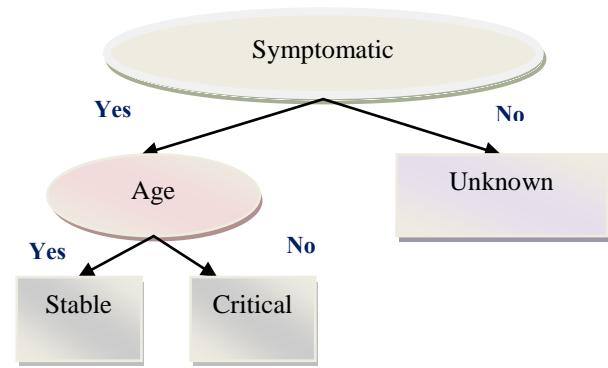


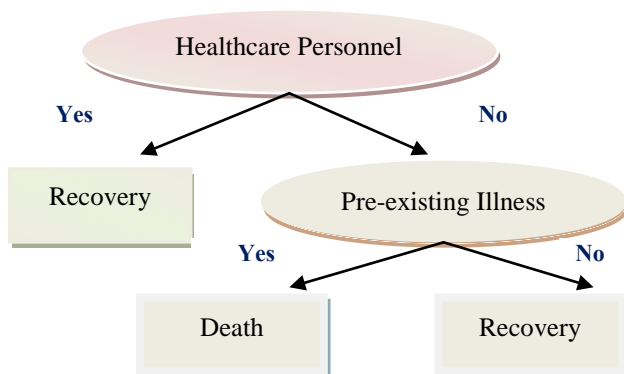
Figure 1.4 J48 Decision Tree Stability Model

II. MERS-COV INFECTION

Nowadays, the outbreak of MERS-CoV attacks caused global attention to Saudi Arabia. The novel virus belongs to the coronaviruses household, which is responsible for creating mild to moderate colds. The control and command center of Saudi Ministry of Health problems a regular report on MERS-CoV contamination cases. The contamination with MERS-CoV can cause critical troubles, nevertheless little information is famous about this novel virus. Two data mining techniques in order to better understand the stability and the possibility of recovery from MERS-CoV infections are used [1]

In 2012, Saudi Arabia seen the episode of a virus named Heart East Respiratory Syndrome Coronavirus (MERS-CoV). The virus belongs to the coronaviruses household which will be responsible for creating moderate to moderate colds. MERS-Co is blamed for creating extreme acute respiratory illness that causes demise in many cases. MERS-CoV signs contain: cough, fever, nose obstruction, air shortness, and sometimes diarrhoea. The virus began distributing quickly in Saudi Arabia in 2013. Ever since then, the Get a handle on and Command Centre of Saudi Ministry of Health in Saudi Arabia started saving and reporting the cases. The ministry internet site offers daily statistics on new established MERS-CoV cases, recoveries, and deaths. Infection with MERS-CoV can result in fatal complications. Unfortunately, there is small information about how a virus advances and how individuals are affected. Knowledge mining is the exploration of large datasets to extract concealed and formerly unknown patterns and relationships [18].

In healthcare, data mining practices have already been widely applied in different programs including: modelling health outcomes and predicting patient outcomes, evaluation of therapy performance, hospital position, and disease control. Several ways are here to predict the security of the event and the chance of recovery from MERS-CoV infection. The goal is to better realize which factors donate to complications of the infection. The types are made by applying knowledge mining practices to the information provided by the Get a handle on and Command Centre of Saudi Ministry of wellness website. Middle East respiratory problem corona virus (MERS-CoV) was isolated in September 2012 from a 60-year-old man from Bisha , Saudi Arabia .In July 2012, that catalos case-patient was hospitalized for severe respiratory condition in Jeddah at Dr. Soliman Fakeeh Clinic and eventually died .As of December 3, 2014, a total of 897



laboratory-confirmed cases of MERS-CoV illness, including 325 deaths, had been described to the World Health Organization; >85% of described MERS-CoV cases and deaths have occurred in Saudi Arabia. The clinical problem among hospitalized MERS-CoV patients involves severe intense respiratory condition, sometimes associated with hypoxemic respiratory distress and extra pulmonary organ dysfunction; however, milder condition and asymptomatic attacks have already been identified through contact investigations [1]. Indication of MERS-CoV to health treatment workers (HCWs) has been described, though no sustained neighbourhood transmission has been identified. A zoonotic source of MERS-CoV has been hypothesized; camels probably may play a role in transmission though the precise types of exposure associated with preliminary cases remain unknown. There is a scarcity of data on how MERS-CoV is spread and on transmission dangers to HCW and other close contacts. Our objectives were to gauge the amount and character of HCW contact with the MERS-CoV case-patient and to serologically assess HCWs for MERS-CoV infection. The subject research was conducted in April 2012; we anticipated development and validation of MERS-CoV serologic assays before completing the study [18].

a. Symptoms of MERS-CoV

Symptoms of Infection with The Middle East Respiratory Syndrome-Coronavirus (MERC-CoV). In line with the found cases up to now, the outward symptoms may possibly contain the next:

- Fever and cough
- Shortness of breath
- Congestion in the nose and throat
- Diarrhoea
- In advanced cases, the in-patient can have really significant difficulties, which can result in demise, such as for example: severe pneumonia.

b. Transmission of MERS-CoV

Based on the confined data accessible up to now, there's number exact clinical evidence distinguishing the way in which through that the disease may be transported from anyone to another. But, it's apt to be like the different types of Corona virus. MOH is creating every work, in cooperation using its companions in the global agencies, like the WHO, to find out about methods for indication, including through animals. The methods for human-to-human Corona indication contain the next:

- Primary indication through drops made all through coughing or sneezing.
- Oblique indication through pressing areas and units contaminated with the disease, and then pressing the mouth, nose or eyes.
- Primary experience of contaminated patients.
- Placed on face-masks just if you're ill or visiting ill patients.

c. Ways of protection against The Middle East Respiratory Syndrome-Corona virus (MERC-CoV)

The Ministry of Wellness suggests people and citizens usually to adhere to these wellness recommendations

to restrict the episode of virus and contagious respiratory conditions till we all know the means of the disease sign:

- Rinse the hands properly and continuously with water and soap, or with different disinfectants employed for give cleaning, especially following coughing, sneezing and applying toilets.
- Decide to try around probable in order to avoid pressing the eyes, nose and mouth together with your give, since it may send the disease following pressing materials contaminated with the virus.
- Prevent around probable calling with contaminated people.
- Use handkerchief when coughing or sneezing and protect the mouth area and nose with it, then remove the handkerchief in a spend holder and ensure that you rinse your hands. If you can find number handkerchiefs, it's chosen to cough or sneeze in the most effective of one's supply, maybe not in the fingers (palms).
- Wear face-masks just if you're ill or visiting ill individuals.
- Stick to hygiene.
- Focus on different wellness behaviours because the natural stability and physical exercise, along with getting enough reveal of rest, since it assists to improve your body's immunity.

III. SCOPE OF COMPARITIVE STUDY

Mers cov disease continues to be an open area of research in medical research and it is discovered to be difficult job to detect it. Scope of comparative examine is to enhance precision of mers cov disease predictive models. This paper has presented an evaluation on numerous predictive model algorithms or technique. The review has clearly revealed that each technique has a unique advantages and restrictions around each other. Here is separated the review in to various disease detecting methods where each technique has a unique advantages and limitations. Moreover comparison of various research papers in the phrases of technique applied, dilemmas mentioned, and parameters based have been revealed in the form of table. These algorithms are:

J48 Algorithm: J48 is a decision tree that is implementation of ID3 algorithm. If we have a dataset in which there are predicators list, targets dependent or independent variables then j48 decision tree will allow you to find out the target variable automatically and make out tree of dataset also. This technique useful for predicting out categorical outcomes. This algorithm uses greedy iterative approach. The algorithm surfaces the dataset based on the most effective educational attribute. At each iteration, the attribute with maximum obtain ratio is picked since the splitting attribute. J48 algorithm has several advantages. They are simple to read and are proven to have comparable reliability to different classification models.

Naive bayes Algorithm: It is really a probabilistic model based on bayes theorem. It assumes type conditional liberty, where in actuality the dependencies between type qualities are ignored. Research has shown that trusting bayes classifiers have equivalent performance to other classification algorithms such as for example decision trees and neural networks. Furthermore, they make highly appropriate models and may

handle big datasets. The Bayesian Classification presents a watched understanding technique as well as a statistical technique for classification. It considers a main probabilistic design and it permits us to record uncertainty about the design in a principled way by determining probabilities of the outcomes. It may solve diagnostic and predictive problems. Bayesian classification offers realistic understanding algorithms and prior understanding and seen information can be combined. Bayesian Classification offers a useful perception for knowledge and considering many understanding algorithms. It figures specific probabilities also.

AdaBoostMI Algorithm:- It is Adaptive Boosting that is a machine learning meta-algorithm. It can be utilized along with several different types of understanding formulas to boost their performance. The production of the other understanding formulas is combined in to a weighted sum that shows the last production of the increased classifier. AdaBoost is versatile in the sense that subsequent weak learners are modified in favour of those cases misclassified by prior classifiers. AdaBoost is painful and sensitive to noisy information and outliers. In some issues it can be less vulnerable to the over fitting issue than different understanding algorithms. The person learners may be weak, but the performance of each one is somewhat much better than random guess.

LogitBoost Algorithm:- In the logistic variant, the LogitBoost algorithm is used to create an LR design at every node in the tree; the node is then separate using the C4.5 criterion. Each LogitBoost invocation is warm-started from their effects in the parent node. Ultimately, the tree is pruned. The fundamental LMT induction algorithm uses cross-validation to get several LogitBoost iterations that doesn't over fit working out data. A faster version has been planned that uses the Akanke data criterion to control LogitBoost stopping.

Part Algorithm:- PART is really a separate-and-conquer principle learner. The algorithm making pieces of principles called choice lists which are in the pipeline pair of rules. A new data is compared to each principle in the list consequently, and that is assigned the type of the initial corresponding rule. PART forms a partial C4.5 choice tree in each time and makes the best leaf in to a rule. PART is really a partial choice tree algorithm, which is the created edition of C4.5 and RIPPER algorithms. The main specialty of the PART algorithm is that it does not require doing worldwide optimisation like C4.5 and RIPPER to make the right principles. However, choice trees are some time more problematic as a result of larger size of the tree which may be oversized and might perform defectively for classification problems.

One R Algorithm:- OneR is just a very simple, quicker and one-level decision tree algorithm. It chooses one-by-one characteristics from the dataset and produces a different group of rules based on problem rate from the training set. Eventually, it decides the feature that gives rules with minimal problem and constructs the last decision tree. The algorithm sees loads of discrete attributes based on very simple association rules involving just one feature in situation part. It

generally realizes attribute fat and then produces a easy rule because of it and estimate its error.

LMT Algorithm:- In computer technology, a logistic model tree (LMT) is just a classification design having an associated watched teaching algorithm that combines logistic regression (LR) and decision tree learning. Logistic design woods are based on the early in the day idea of a style tree: a determination tree that's linear regression versions at their leaves to provide a piecewise linear regression design (where ordinary decision woods with constants at their leaves could produce a piecewise regular model). Classifier for making logistic model woods, which are classification woods with logistic regression operates at the leaves. The algorithm may deal with binary and multi-class goal factors, numeric and small characteristics and missing values.

JRip Algorithm:- This algorithm introduces a propositional rule learner, Repeated incremental pruning to produce error reduction. It has some stages as Building stage, Grow phase, Prune phase, Optimization stage. JRip (RIPPER) is among the standard and hottest algorithms. Courses are analyzed in raising measurement and an original group of principles for the type is created applying slow decreased mistake JRip (RIPPER) profits by managing most of the types of a certain judgment in the instruction information as a type, and locating some principles that protect most of the customers of this class. Afterwards it profits to another type and does the exact same, saying that till all courses have been covered.

DecisionStump Algorithm:- A decision stump is a device understanding product consisting of a one-level choice tree. That's, it's a choice pine with one central node (the root) that is straight away linked to the final nodes (its leaves). A determination stump makes a forecast on the basis of the price of merely a simple feedback feature. Often they're also known as 1-rules. With respect to the form of the feedback function, many modifications are possible. For small characteristics, it's possible to construct a stump which has a leaf for every probable function price or even a stump with both leaves, certainly one of which fits for some picked type, and one other leaf to the rest of the categories. For binary characteristics those two systems are identical.

Bagging Algorithm:- Bootstrap aggregating, also referred to as bagging, is a device understanding collection meta-algorithm developed to boost the balance and reliability of equipment understanding methods utilized in mathematical classification and regression. In addition it decreases deviation and assists in order to avoid over fitting. Though it is normally put on choice pine techniques, it can be utilized with any kind of method. Bagging is just a particular event of the design averaging approach. It is the way in which reduce the deviation of one's forecast by generating extra information for instruction from your own unique dataset applying mixtures with representatives to create multiset of exactly the same cardinality/size as your unique data. By rising how big is your instruction collection you cannot increase the design predictive power, but only reduce the deviation, narrowly focusing the forecast to estimated outcome.

Stacking Algorithm:- Stacking is a procedure where in fact the result of 1 degree of classifiers is employed as insight for another level. That's, the forecasts of some classifiers would be

the functions for different classifiers. Because of this, you will have to train among the versions with the results of the initial classifier as input. Which among the classifiers must apply wherever depends in your certain application. Likewise, how to complete this will depend on which process you have applied to teach these classifiers. It is really a much like increasing: additionally you use a few versions to your unique data. The huge difference here's, but, that there isn't only an scientific system for the fat purpose, somewhat you present a meta-level and use yet another model/approach to calculate the insight as well as results of each and every design to calculate the loads or, quite simply, to find out what versions accomplish effectively and what defectively provided these insight data.

IV. LITERATURE REVIEW

Isra Al-Turaiki *et al.* (2016) [1] introduced the episode of MERS-CoV attacks that triggered worldwide attention to Saudi Arabia. The disease with MERS-CoV can result in dangerous issues, nevertheless small information is famous about this story virus. In this paper there is applied two data mining techniques to be able to better realize the stability and the likelihood of recovery from MERS-CoV infections. Naive Bayes classifier and J48 choice pine algorithm were used to build our models. The dataset used includes 1082 documents of instances noted between 2013 and 2015. As for the stability versions using J48, two attributes were discovered to be essential for predicting stability: symptomatic and age. Previous individuals have reached large risk of developing MERS-CoV complications. Ultimately, the performance of all the versions was examined using three steps: reliability, accuracy, and recall. In general, the reliability of the versions is between 53.6% and 71.58 percent. Data mining has been generally employed for the prognosis and diagnoses of numerous diseases. Ferreira *et al.* [8] used data mining to increase the diagnosis of neonatal jaundice in newborns. In their experiment, the dataset contained 70 variable collected for 227 balanced newborns. Several information mining practices were used, including: J48, CART, Trusting Bayes classifier, multilayer perception, SMO and easy logistic. The most effective predictive models were acquired by utilizing Trusting Bayes, multilayer perception, and easy logistic.

For center disease diagnoses, Venkatalakshmi and Shivsankar [9] compared the efficiency of choice pine algorithm and Trusting Bayes. The fresh benefits applying a dataset of 294 documents with 13 features showed that the efficiency of the 2 calculations is com-parable. FP-growth, Association principle mining, and decision woods were employed for the examination and prognosis of chest cancer [3]. The classification models were created employing a dataset of 699 documents and 9 attributes and the very best reliability was reached using decision woods induction algorithms. With regards to survivability predicting, Bellaachia *et al.* [10] applied Trusting Bayes, back-propagated neural system, and the C4.5 choice pine algorithm to predict the survivability of chest cancer patients. The dataset found in the research was acquired from the Surveillance Epidemiology and Conclusion Benefits (SEER). Experimental benefits suggested that the C4.5 algorithm out performed another two techniques. Recently, a few predictive versions for chest cancer success were produced [4]. The versions were based on a dataset of 657,712 files and 72 variables, also acquired from

SEER. Three different data mining techniques were applied: Help Vector Machine (SVM), Bayes Systems, and Chisquared Automatic Connection Recognition (CHAID). Results showed that the very best success forecast design was obtained applying SVM. The writers in [4] showed a study of predictive versions for chest cancer survival. The key aim was to find important attributes that donate to chest cancer survival. Three data mining techniques were applied: Support Vector Unit (SVM), Bayes Internet, and Chisquared Automatic Connection Recognition (CHAID). Experiments on a dataset acquired from SEER showed that the SVM design outperformed different models in phrases of reliability, tenderness, and specificity. SVM surely could recognize twenty characteristics that are important indications of chest cancer survivability.

Sandhu *et al.* [5] planned a cloud-based MERS-CoV prediction system. The device is dependent on Bayesian Belief Systems (BBN) for original classification of patients. A geographical placing process is utilized to signify people on Bing Maps. People classified as contaminated were monitored applying GPS from their mobile phones. The planned process is helpful to citizens as it enables them to prevent contaminated areas. In supplement, healthcare authorities may manage the illness issue more effectively. The BBN achieved an reliability of 83.1% on artificial data. Memish ZA *et al.* [11] considered that A zoon tic source of MERS-CoV has been hypothesized; camels probably may play a role in sign though the particular kinds of publicity related to primary instances stay unknown. There stays a scarcity of data on how MERS-CoV is distributed and on sign dangers to HCW and other shut contacts. Our objectives were to judge their education and character of HCW connection with the MERS-CoV index case-patient and to serologically evaluate HCWs for MERS-CoV infection. The subject research was conducted in March 2012; and anticipated growth and validation of MERS-CoV serologic assays before performing the study. Uzma Faridi *et al.* [12] described that Centre East respiratory problem could be the intense respiratory problem brought on by betacoronavirus MERS-CoV. The very first situation of the infection was noted from Saudi Arabia in 2012. That disease is life-threatening and is really a shut general of a serious intense respiratory problem (SARS) that is responsible for significantly more than 3000 deaths in 2002–2003. Based on Ministry of Wellness, Saudi Arabia. Due to the understanding and early examination the likelihood is slipping gradually. MERS-CoV and SAR-CoV are from the exact same genus, therefore it had been believed that the medications which prevent the development of SARS-CoV also can prevent the development of MERS-CoV but these medications aren't entirely inhibiting disease activity. Till here is not appropriate design and treating MERS-CoV.

Hyuk-Jun Chang *et al.* [2] introduced opinion of the fundamental imitation number of the coronavirus (CoV) of MERS on the basis of the noted information from the MERS spread in South Korea. To the conclusion we utilize a mathematical product defined by a couple of standard differential equations, *i.e.* the well known susceptible-infected-removed (SIR) model. First we match the product to the epidemic contour information obtained from the outbreak. Then we are able to identify the product variables and also the fundamental imitation number. Observe that there had been no get a grip on treatment throughout the first period of the MERS outbreak in South Korea, which may be considered as the most effective issue for the opinion examine of intrinsic

epidemic variables of MERS, such as for instance standard imitation number. Shweta Kharya et al. [6] introduced that Breast cancer is among the primary cancers for females in produced nations including India. It's the 2nd most frequent reason behind cancer demise in women. The large likelihood of chest cancer in girls has improved considerably within the last years. In that report we've mentioned various data mining approaches which have been applied for chest cancer

examination and prognosis. Breast Cancer Examination is unique of benign from malignant Breast mounds and Breast Cancer Treatment anticipates when Breast Cancer is always to recur in individuals which have had their cancers excised. Recent study being moved out applying the data mining techniques to boost the breast cancer examination and prognosis.

V. COMPARISON TABLE

Table 1 has clearly shown the contrast among the available mers cov disease detecting methods. Though the contrast has

clearly shown that the J48 algorithm has better effects compared to the accessible method.

Table 1: Comparative analysis

Reference no.	Title	Objective	Technique	Accuracy	Precision	Recall	F-Measure	Re-production no.
[1]	Building predictive models for MERS-CoV infections using data mining techniques	To better understand the stability and the possibility of recovery from MERS-CoV infections.	Naive Bayes classifier and J48 decision tree algorithm	✓	✓	✓	×	×
[2]	Evaluation of the basic reproduction number of MERS-CoV during the 2015 Outbreak in South Korea	An estimation of the basic reproduction number of the coronavirus (CoV) of MERS	Employed a mathematical model described by a set of ordinary differential equations	×	×	×	×	✓
[3]	Data Mining Techniques For Diagnosis And Prognosis Of Cancer	Diagnosis and prognosis of cancer	FP algorithm in Association Rule Mining (ARM)	×	×	×	✓	×
[4]	Prediction of Breast Cancer Survival Through Knowledge Discovery in Databases	To develop predictive models and discover relationships between certain predictor variables and survival in the context of breast cancer	IBM SPSS Statistics 16, Access 2003 and Excel 2003 ,IBM SPSS Modeller 14.2 , Support Vector Machine (SVM) model	✓	✓	×	×	×
[5]	An intelligent system for predicting and preventing MERS-CoV infection outbreak	Predicts MERS-CoV-infected patients	Bayesian belief network	✓	✓	×	×	×
[6]	Using Data Mining Techniques for diagnosis and prognosis of cancer disease	To enhance the breast cancer diagnosis and prognosis	Neural Network, Association Rule Mining,, Naive.Bayes,C4.5 decision tree algorithm,	✓	×	✓	×	×

			Bayesian Networks					
[7]	Survey on Predictive Analysis of Diabetes in Young and Old Patients	Recognizes possibility of diabetes for particular person	Classification Technique, Bayesian network, Decision Induction Algorithm, R Studio	✓	×	×	✓	×
[13]	A novel behavioural model of the pasture-based dairy cow from GPS data using data mining and machine learning techniques	To build a robust model of the most notable behaviours of the pasture based dairy cow: grazing, resting, and walking and to fully document how this was derived.	JRip , Random forest, SEM	✓	✓	×	✓	×
[14]	Wheat grain classification by using dense SIFT features with SVM classifier	An extensive Experimental study using a smart decision system with SVM and histogram features is conducted on given dataset.	Support Vector Machine (SVM) classifier, k-means clustering	✓	✓	×	×	×
[15]	Diagnosis of Breast Cancer using Decision Tree Data Mining Technique	Summarize Breast cancer , Demonstrate Decision Tree ,	Decision trees, j48 algorithm	✓	✓	✓	×	×

VI. CONCLUSION

In that report, there is developed many designs to predict the security and healing of mers-cov infections. here is given out designs that were developed applying trusting bayes and j48decision pine classification algorithms. the decision tree healing product suggested that people who are healthcare workers are more prone to survive. this feature was discovered to be crucial in predicting the security of the patient. previous patients with ages between 66 and 87 are more prone to suffer from important complications. the efficiency of all the designs was considered and compared. in general, the reliability of the designs is between 53.6%and 71.58%.it is genuinely believed that the efficiency of the forecast designs could be improved with the use of more individual data. as potential perform, here is plan to right contact hospitals in riyadh so as toc collect more info linked to people with mers-cov infections.

VII. REFERENCES

- [1] Al-Turaiki, Isra, Mona Alshahrani, and Tahani Almutairi. "Building predictive models for MERS-CoV infections using data mining techniques." *Journal of Infection and Public Health* 9.6 (2016): 744-748.
- [2] Chang, Hyuk-Jun. "Evaluation of the basic reproduction number of MERS-CoV during the 2015 outbreak in South Korea." *Control, Automation and Systems (ICCAS), 2016 16th International Conference on*. IEEE, 2016.
- [3] Majali, Jaimini, et al. "Data mining techniques for diagnosis and prognosis of cancer." *International Journal of Advanced Research in Computer and Communication Engineering* 4.3 (2015): 613-6.
- [4] Afshar, Hadi Lotfnezhad, et al. "Prediction of breast cancer survival through knowledge discovery in databases." *Global journal of health science* 7.4 (2015): 392.

- [5] Sandhu, Rajinder, Sandeep K. Sood, and Gurpreet Kaur. "An intelligent system for predicting and preventing MERS-CoV infection outbreak." *The Journal of Supercomputing* 72.8 (2016): 3033-3056.
- [6] Kharya, Shweta. "Using data mining techniques for diagnosis and prognosis of cancer disease." arXiv preprint arXiv:1205.1923 (2012).
- [7] Komal Agicha, Priyanka Bhatia, Neha Badlani, Ashutosh Menghrajani, Abha Tewari, "Survey on Predictive Analysis of Diabetes in Young and Old Patients"
International Journal of Advanced Research in Computer Science and Software Engineering Volume 5, Issue 10, October-2015
- [8] Ferreira, Duarte, Abílio Oliveira, and Alberto Freitas. "Applying data mining techniques to improve diagnosis in neonatal jaundice." *BMC medical informatics and decision making* 12.1 (2012): 143.
- [9] Venkatalakshmi B, Shivsankar M. "Heart disease diagnosis using predictive data mining", *International Journal of Innovative Research in Science. Engineering and Technology* 2014;3:1873—7.
- [10] Bellaachia A, Guven E. "Predicting breast cancer survivability using data mining techniques", in: *Ninth Workshop on Mining Scientific and Engineering Datasets in conjunction with the Sixth SIAM International Conference on Data Mining*; 2006.
- [11] Memish, Ziad A., et al. "Middle East respiratory syndrome coronavirus in bats, Saudi Arabia." *Emerg Infect Dis* 19.11 (2013).
- [12] Faridi, Uzma. "Middle East respiratory syndrome coronavirus (MERS-CoV): Impact on Saudi Arabia, 2015." *Saudi Journal of Biological Sciences* (2016).
- [13] Williams, M. L., et al. "A novel behavioral model of the pasture-based dairy cow from GPS data using data mining and machine learning techniques." *Journal of dairy science* 99.3 (2016): 2063-2075.
- [14] Olgun, Murat, et al. "Wheat grain classification by using dense SIFT features with SVM classifier." *Computers and Electronics in Agriculture* 122 (2016): 185-190.
- [15] Sumbaly, Ronak, N. Vishnusri, and S. Jeyalatha. "Diagnosis of breast cancer using decision tree data mining technique." *International Journal of Computer Applications* 98.10 (2014).