Abstract—Large volume of data about patients and their clinical information is present in medical databases. For extracting the features and their connections from a tremendous database, different data mining methods should be utilized. As Liver transplantation is the corrective surgery for the patients experiencing end-stage liver disease, anticipating the survival rate after Liver transplantation has a major effect. Suitable determination of attributes and strategies are fundamental for the survival prediction. The system proposed cloud-based hybrid method for prediction of long-term survival of liver transplantation patients. From 389 attributes of the United Nations Organ sharing registry, data with 256 attributes were collected for the survival prediction. A tenfold cross validation (CV) was applied in the medical input dataset which was get from United Network Organ Sharing database. So as to play out the dimensionality decrease of an immense database, principal component analysis (PCA) with ranking was done. The relation between attributes was predictable and verifies using different association rule mining methods, such as Apriori algorithm. We evaluated the rules generated by the association rules mining algorithm before and after PCA was also executed, for verifying the results. Proposed an effective and accurate artificial neural network (ANN) model for the prediction of long-term survival of liver patients who undergo liver transplantation (LT) and then predicted data is uploaded on the AWS cloud.

Keywords—Cloud based Hybrid classifier, Liver Transplantation (LT), Artificial neural network (ANN), ten-fold cross-validation (CV), Amazon Web Services (AWS) cloud.

1. INTRODUCTION

Security and privacy risks, healthcare organizations can certainly take advantage of cloud computing solutions and bring tremendous benefits such as help to improve patient quality of service and reduce overall healthcare costs. Liver transplantation (LT) is a life-saving procedure for patients suffering from chronic and acute liver failure and has undergone massive development. Advances in the inhibition of infectious and non-infectious complications, the stability of immunosuppressive therapy, and the distribution of organs are central features that have developed since the first series of LTs. Long-term survival and quality of life are the main goals of LT. Predicting survival is an important factor used to determine the success of liver transplant surgery. Despite the fact that liver transplantation is the most favorable treatment method for those who suffer from liver diseases, the short age of donor liver sharply limits this option [1] [2]. The number of patients waiting for a liver transplant is high, but the number of people trying to donate a liver is much lower, so the mortality rate is increasing every year. Medical experts, based on the Model of End-stage Liver Disease (MELD) score, have made judgments about the liver transplantation and the prediction of the outcome of the transplant. Of the three parameters in the MELD score of bilirubin, creatinine and the International normalized ratio (INR), creatinine levels vary depending on the weight of the liver recipient. According to the principle of the MELD score, patients at the top of the waiting list are given the highest priority in organ allocation for liver transplantation. But in this case, some patients can get the donor organ early, while others have to wait longer for the donor organ, which leads to a lower survival rate [3].

Instead of using the MELD score, continuous research into more accurate models to predict the long-term survival of liver transplant patients has led to the introduction of more accurate models, such as Artificial neural networks (ANNs). ANN uses computer technology to model systems that structurally and functionally recall biological neural networks: they consist of a set of highly interconnected processing units (neurons) linked to weighted connections and include an input layer, an output layer and one or more hidden layers. The input layer consists of the different data available for analysis and the output layer consists of the different results.

In medical field, several commonly accepted statistical models are used to predict LT, without taking into account the donor-recipient comparison, donor organs are allocated to the first patient in the queue because of the shortage of donor organs in the classical method of allocation. But ANNs play a more important role by avoiding the local optimum of traditional statistical methods and logistic regression models. ANN models are successfully used in traditional statistical methods for predicting survival, taking into account the characteristics of the donor-recipient and transplanted organs. The purpose of this study [4][5] was to develop a Prediction of Long Term Survival after Liver Transplantation system based on matching donor and recipient. There are many reasons for developing this system such as existing selection / distribution systems are based on the risk of patient death in the waiting list and do not recognize differences in the quality of donor organs; efforts to increase the number of donor organs are likely to
result in a relatively high proportion of donors with expanded criteria; comparison of donors and recipients may provide an opportunity to predict results at the time when a particular donor liver will be allocated to a specific recipient; differences in local acceptance rates and policies may be reduced; and Overall outcome and effectiveness may improve.

This paper proposes a cloud-based hybrid classifier for prediction of long-term survival of liver transplant patients. Our system uses 28 attributes from UNOS dataset and performs the feature extraction on dataset by using Principal component analysis (PCA) [6] algorithm, and classifies the data set into training and test sets by using proposed hybrid classifier. The corresponding donor-recipient pairs were selected using ten-fold cross-validation in the training of medical data. The proposed efficient and accurate artificial neural network (ANN) model predicts the long-term survival of liver patients who undergo liver transplantation (LT), and then the predicted data is uploaded to the Amazon Web Services (AWS) cloud. Cloud computing plays a significant job [7] in optimizing resources, reducing the combined charge of healthcare, and sharing a new period of improvement, because of this information is accessed anytime, anywhere, which can be attain when moving healthcare information to the cloud.

II. REVIEW OF LITERATURE

The different analyst’s conduct investigation on ANNs model to forecast patient survival after liver transplantation. Raji et al.[1], effective and accurate model of an ANN for predicting the long-term survival of liver patients undergoing liver transplantation (LT) has been proposed. They performed a 13-year survival analysis in predicting liver patients after LT. The author trained liver observation data for 13 years separately using the multi-layer perceptron (MLP) model with ANN with the correct selection of data attributes. The author [8] Raji et al. introduces the several machine learning methods to predict increased survival after LT. Methods based on artificial neural networks are widely used to study medical data and predict survival outcomes. The role of machine learning tools in the medical field is growing day by day as medical data grows exponentially, where doctors cannot easily identify hidden patterns and useful information occurring in large volumes of data. For performing machine learning operations in engineering, medicine, mathematics, Economics, science, Geology, and many other fields, the role of ANNs is remarkable.

Ayllon et al. [10], a donor-recipient (D-R) matching model for liver transplantation (LT) based on artificial neural networks (ANNs) from a Spanish multicenter study is reported. The goal is to test the ANN-based methodology in another European health system to confirm it. The ANN model was developed using a cohort of patients from King’s College hospital. ANN was trained and tested using KCH pairs for both 3-and 12-month survival models. The endpoints were the probability of graft survival and non-survival. The final model is a rule-based system to facilitate the decision on the most appropriate D-R comparison. Bertocco et al. describes the fact that the liver transplant procedure are common and very expensive, their cost structure is still poorly understood.

Ortiz et al. [11], propose a new algorithm for ordinal classification based on a combination of ensemble and discriminant analysis methods. This proposal relates to the actual application of liver transplantation, where the goal is to predict the survival of the transplant. This proposal was tested using an important Biomedicine problem: predicting graft survival in a real-world data set for liver transplantation. In paper [12] Nair et al. proposes a model for predicting life expectancy after liver transplantation and model is deployed using the automatic learning process. They use Moderate 200 attributes from UNOS dataset. The author performs clustering for efficient clustering between Donor data and Recipient data by using K-mean algorithm. Chandra et al. [13], proposes a MLP model for long term survival of patients after Liver transplantation. They use 256 attributes, out of 389 attributes from UNOS dataset. The PCA is used for feature extraction. The system performance was evaluated with Radial Basis Function (RBF).

III. PROPOSED SYSTEM

A. Proposed System Architecture

Liver transplantation is the ultimate treatment for people who suffer from end-stage liver disease. With various advances in the field of liver transplantation, the survival rate is increasing day by day, and patients are much dependent on this surgical procedure. In the liver transplantation, it is very difficult to obtain the appropriate donor-recipient agreement. The survival of a patient after a liver transplant depends on proper donor recipient matching to predict. The characteristics of receiving a donor, etc., and these characteristics match. The manual method for matching donor-recipient functionality is tedious. In many research subjects, the dimension of the input is too big. This can lead to a relatively slow process of mining. Figure 1 shows the cloud based hybrid classifier for the prediction of long-term survival of liver patients.

Configure neural network using following feature:
1. Neurons =12345
2. 4 hidden layer with different activation layer like TANH,RELU respectively
3. STOCHASTIC GRADIENT DESCENT
4. learningRate = 0.01
5. lstmLayer=2 with each neuron layer contain 256 neuron
6. gateActivationFunction=HARDSIGMOID
7. lossFunction=MSE

1) UNOS Dataset:
To collects the data from the United Nations Organ Sharing (UNOS) dataset. The dataset contains 389 attributes, and of these 389, only 70 attributes consist of donor attributes, transplant attributes, and recipient attributes, and of these 70, only 28 attributes are used. The table I, shows the 28 attributes and variables of donor and receiver.
TABLE I. TABLE OF ATTRIBUTE NAME, VARIABLE TYPE, AND DONOR/RECEIVER

| Attributes Name   | Variable Type | Donor/Receiver |
|------------------|---------------|----------------|
| GENDER_DON       | Nominal       | Donor          |
| NON_HRT_DON      | Nominal       | Recipient      |
| SGOT_DON         | Numeric       | Donor          |
| SGPT_DON         | Numeric       | Donor          |
| TBLII_DON        | Numeric       | Donor          |
| AGE_DON          | Numeric       | Donor          |
| CLIN_INFECT_DON  | Nominal       | Donor          |
| CREAT_DON        | Numeric       | Donor          |
| DIABETES_DON     | Nominal       | Donor          |
| DON_TY           | Nominal       | Donor          |
| GENDER           | Numeric       | Receiver       |
| INIT_AGE         | Numeric       | Receiver       |
| MALIG_TCR        | Nominal       | Receiver       |
| FINAL_BILIRUBIN  | Numeric       | Receiver       |
| FINAL_INR        | Numeric       | Receiver       |
| FINAL_MELD_OR_PELD | Nominal  | Receiver       |
| FINAL_MELD_PELD_LAB_SCORE | Numeric | Receiver       |
| FINAL_SERUM_CREAT | Numeric   | Receiver       |
| FINAL_SERUM_SODIUM | Numeric  | Receiver       |
| BMI_TCR          | Numeric       | Receiver       |
| ENCEPH_TCR       | Nominal       | Receiver       |
| EXC_HCC          | Nominal       | Receiver       |
| FINAL_ALBUMIN    | Numeric       | Receiver       |
| ALKPHOS          | Numeric       | Receiver       |
| FINAL_ASCITES    | Numeric       | Receiver       |
| NUM_PREV_TX      | Numeric       | Transplantation|
| TX_LIV           | Nominal       | Transplantation|

2) Data Pre-processing:
In this data preprocessing stage unwanted data or irrelevant data is removed. UNOS dataset contains 389 attributes form this only 256 attributes are related to liver transplantation patients. For mining 27 attributes used PCA to dimensionality reduction. According to ranking with standard deviation, here consider 27 relevant attributes which are very useful for survival prediction after LT. The RNN and MLP algorithm is used to train neural networks and this method approximates the error of the networks within a second order expression which contrasts to the back-propagation algorithm that does it with the first-order expression.

3) Hybrid Classifier
The system performs MLP, RNN and Hybrid classifier. Here, 28 attributes input is given to classifiers. The hybrid classifier model trained using the clinical attributes of LT patients and training to classifiers is used feedforward algorithm.

The input data consist of donor data and transplant attributes which are AGE DON, CLIN INFECT DON, CREAT DON, DIABETES DON, DON_TY, GENDER DON, NON HRT DON, SGOT DON, SGPT DON, TBLII DON, BMI TCR, ENCEPH TCR, EXC HCC, FINAL ALBUMIN, FINAL ASCITES, FINAL BILIRUBIN, FINAL INR, FINAL MELD OR PELD, FINAL SERUM SODIUM, GENDER, INIT AGE, MALIG TCR, GSTATUS, NUM PREV TX, TX LIV, FINAL MELD PELD LAB SCORE, and FINAL SERUM CREAT.

In the hidden layers, the sigmoid function is used to training the clinical data. For the best survival of patients which is undergoing LT, proper matching of donor-recipient is required and this is done because of used hidden layers during training.

Calculate hidden layer Y formula such as:

\[ y = \frac{t}{2} \]  

eq. (1)

Where, \( t = I + O \);

 затем, \( Y = (I+O)/2; \)

Here, I denote the number of attributes and Classifier O denotes the number of classes.

4) Prediction of Liver Transplantation:
The survival rate of LT is the time between the occurrence of life and death. The survival analysis of each liver patient has been carried out for several years so far.

B. Algorithm
1) Algorithm 1: Hybrid classifier Algorithm
1. Take input dataset
2. Take each entry of dataset and convert it into matrix form
3. Assign randomly weight and bias matrix
4. Do multiplication of weight, bias and input matrix for each hidden layer
5. Generate result matrix
6. Repeat step 1 and 5 for each entry in the dataset.
2) Algorithm 2: Pseudo Code for Levenberg-Marquardt (LM)-MLP:

Forward pass:

steps 1: take input and get initial weights of input layer...
steps 2: calculate output of input layer using input value and weights...
steps 3: take output of input layer as input to the hidden layer...
step 4: calculate output of hidden layer using input of hidden layer...
step 5: take output hidden layer as input to the output layer.
step 6: calculate output of output layer.
step 7: calculate total error.

Backward pass:

step 8: Backpropagate from the inputs of the output layer to the outputs of the hidden layer
step 9: Backpropagate from the outputs of the hidden layer to the inputs of the hidden layer
steps 10: Backpropagate from the inputs of the hidden layer to the outputs of the input layer
step 11: Backpropagate from the outputs of the input layer to the inputs of the input layer

After performing forward pass and backward pass, calculate jacobian matrix by using updation rule, total error goes down.

C. Mathematical Model

Let ‘S’ be the system that predicts survival of patient after LT using machine learning, S = { I, O, F}

I = { Input to train the system } O = { Output of the system }

F = {functions of the system }

I = { I1 }

I1 = {LT patients dataset with relevant attributes after applying PCA from UNOS }

F = {Apply PCA(), MLP(), RNN(), Hybrid Classifier () Predict()}

- Apply PCA() = { Performs PCA Analysis }
  \[ V^{-1}CV = D \] eq. (2)

where,
  V = Matrix V of eigenvectors
  C = Covariance matrix
  D = Diagonal matrix of eigenvalues of C.

- Initiate classifiers
  - MLP() = {Initiate MLP classifiers on data, Learning rate 0.3, }
  - RNN () = { Initiate RNN classifiers on data, Learning rate 0.02. }
  - Hybrid Classifier()={ Initiate hybrid classifier on data, Learning rate 0.01. }

- Predict() = { Predict survival of the patients }  

2.  \( O = \{ O1 \} \)

- O1 = { Prediction: predicts survival of patient after LT: 0-Survived, 1-Not survived }

Time complexity is usually expressed as a function of the “size” of the problem. The complexity of NN algorithm’s is \( O(n^2) \), where \( n \) is the size of the array to be sorted. Suppose the input vectors have dimension \( n \) by \( 1 \). Then the time it takes to produce the output is a constant multiplication of \( n \) (when we use the big \( O \)). We keep in mind here that all we need is a two-layer neural network. So the time complexity is \( O(n) \) where \( n \) is the size of the input vector.

IV. RESULT AND DISCUSSION

A. Experimental Setup

The system utilizes the Java framework of the Windows platform. The Net bean IDE is used as a development tool. The system does not require you to run specific hardware; any standard machine can run the application. As you acknowledge that you may experience time in the actual cloud. For data sets, the system uses an Amazon Web Services (AWS) cloud.

B. Result

The medical prediction is the evaluation of the cure, the functional level of the organ, the recurrence of the affected disease, complications, period for which the patient is hospitalized, the duration of life of the patient or patient group. Medical expectations help clinical researchers and physicians to discover certain patterns of disease. It also supports medical researchers to allocate resources according to requirements. The prognosis helps to make a decision about the state of the disease. LT has been widely accepted as a therapeutic treatment over the past 50 years. The number of people with liver diseases is preserved by this important medical activity. Medical researchers estimated that the overall result of LT is 88%. Clinically, the prediction of patient survival after LT is based on the End-stage Liver Disease (MELD) score. There are many positive results in the prediction by the MELD score, but there are some disadvantages. The MELD score will not accurately predict short-term mortality in 15%–18% of the patients with chronic liver disease listed for liver transplantation.

The paper uses the LT dataset from UNOS, a promising dataset that has been certified. The formatting of the data set is very important before the operation. The extraction of related attributes from a large data set requires data mining techniques. Extracting related attributes from a medical dataset requires careful attention. The UNOS dataset contains 389 attributes, form this only 256 attributes are related to liver transplantation patients. So we can remove 59 attributes from 256 attributes during preprocessing step. Among the 197 attributes, we performed a PCA to reduce the dimension. According to the standard deviation ranking, we obtained 70 relevant attributes and only 28 attributes are used that are very useful for predicting survival after LT. By generating rules using Association rule mining algorithms, we could prove that our extracted
28 attributes are more suitable for predicting survival in LT. We ranked all the input attributes used to train the model using the WEKA library. Using weight coefficients, we found that all the attributes of the donor, recipient, and transplant have the same significance in ranking. Table II shows the ranking of input attributes after PCA.

| Rank | Score | Attributes          |
|------|-------|---------------------|
| 1    | 89.9  | AGE_DON             |
| 2    | 88.6  | BMI_TCR             |
| 3    | 87.1  | GENDER_DON          |
| 4    | 85.3  | FINAL_ALBUMIN       |
| 5    | 83.5  | CREAT_DON           |
| 6    | 81    | SGOT_DON            |
| 7    | 78.2  | SGPT_DON            |
| 8    | 76.9  | CLIN_INFECT_DON     |
| 9    | 75.3  | FINAL_BILIRUBIN     |
| 10   | 72.1  | FINAL_ASCITES       |
| 11   | 70    | FINAL_INR           |
| 12   | 69.4  | GENDER              |
| 13   | 67.8  | DON_TY              |
| 14   | 67.5  | DIABETES_DON        |
| 15   | 66.6  | INIT_AGE            |
| 16   | 64.2  | TBILI_DON           |
| 17   | 63.8  | EXC_HCC             |
| 18   | 61.9  | FINAL_MELD_PELD_LAB_SCORE |
| 19   | 60    | FINAL_MELD_OR_PELD  |
| 20   | 58.4  | FINAL_SERUM_CREAT   |
| 21   | 58.1  | FINAL_SERUM_SODIUM  |
| 22   | 57.2  | NON_HRT_DON         |
| 23   | 55.9  | ENCEPH_TCR          |
| 24   | 54.7  | MALIG_TRR           |
| 25   | 53.5  | MALIG_TCR           |
| 26   | 51.6  | NUM_PREV_TX         |
| 27   | 50.5  | TX_LIV              |
| 28   | 66.6  | ALKPHOS             |

In this way, the most relevant top-rank attributes are successfully extracted from the dataset using PCA mining algorithms, ranking, and Association rules. The associated attributes from the Association rules were selected to build the prediction model. For any analysis and experiments, the choice of the appropriate model and data set affects the outcome of the event. Although a rich set of data is collected and preserved, only a small subset was used to predict patient survival.

Our proposed framework utilizes distinctive classification algorithms to train more different classifiers so as to make better classification. The essential thought is to utilize a blend of classifiers to normally diminish change and utilize a more grounded algorithm to expressly build classification execution. The MLP and LM are two basically diverse classification algorithms, and so it is appealing to observe that hybrid classification created with them together can attain better classification performance compared to non-hybrid classifier created with them individually. The hybrid classifier (MLP+LM) performs better not in a deterministic but a probabilistic manner. In a hybrid classifier where two classification algorithms are utilized, if one is dissimilar and unbalanced while the other is more precise, then there is a higher probability that we can have a higher value of accuracy gain; the gain is measured against, where only one of the two classification algorithms MLP and RNN is used. The hybrid classifier calculation is done by following formula:

\[ w_{k+1} = w_k - (J^T J + \lambda I)^{-1} J e_k \]  

where,
\( w = \) weight
\( J = \) Jacobian matrix
\( J e = \) Error gradient
\( J^T J = \) cross-product Jacobian

This paper proved that Hybrid model is a powerful for forecasting purposes. We used this model to predict the best outcome of patients after LT. We also experiment our model with existing work and proved that the highest accuracy has been obtained in the hybrid model with the data set. In experiment, accuracy parameter used to evaluate the effectiveness of the Hybrid model. The Hybrid classifier, MLP, and RNN are used to prove the accuracy of proposed cloud based hybrid model. Here True Positive (TP), False Negative (FN), False Positive (FP) and True Negative (TN) terms are used for calculating the accuracy of model. In which, if LT is a failure in the patient, but it implies success, and the result is truly negative (TN). Both a true positive and a true negative proved condition (also called the truth standard). If the diagnostic result shows that the LT is successful, who actually has it fail, and the test result is false Positive (FP). Similarly, if the result of a diagnostic test suggests that the LT is unsuccessful, but it was successfully performed, the test result is false negative (FN). Both false-positive and false-negative test results indicate that they are the opposite of the actual state.

Table III shows that accuracy comparison between hybrid classifier, RNN and MLP algorithm. The figure 2, shows that the Hybrid classifier is more accurate than the MLP and RNN. It shows higher 96.44 % accuracy.

| Algorithm       | Accuracy in % |
|-----------------|---------------|
| Hybrid Classifier | 96.44         |
| MLP             | 89.78         |
| RNN             | 93.32         |

![Fig 2. Accuracy Graph](image_url)
V. CONCLUSION

For the final stage of liver disease the Liver transplantation is the most excellent and most precise solution. The most significant features of liver transplantation are organ collection and sharing. Recently, liver transplantation is one of the demanding areas in the field of organ transplantation. Medical experts predict the survival of patients after liver transplantation based on the MELD score. As creatinine varies depending on the body weight of a liver patient and is lower in women than in men, the research team presented ANN models to predict patient survival after liver transplantation. The system uses UNOS dataset for prediction of Long Term Survival after Liver Transplantation. The predicted data is uploaded onto the AWS cloud. Cloud computing plays an important role in reducing the combined cost of healthcare, optimizing resources, and sharing a new era of innovation, because of this information is accesses anytime, anywhere, which can be achieved when moving healthcare information to the cloud. The system records 96.44% accuracy for survival prediction. We compared the proposed hybrid model with existing MLP and RNN classifiers. Given all this data, for predicting long term survival of patients after liver transplantation, we could estimate that cloud based hybrid model is the most appropriate ANN model.

REFERENCES

[1] C. G. Raji and S. S. Vinod Chandra, “Long-Term Forecasting the Survival in Liver Transplantation Using Multilayer Perception Networks”, International Journal of Innovative Science and Research Technology, Volume 4, Issue 5, May–2019.

[2] Lena Griebel, Hans-Ulrich Prokosch, Felix Kopcke, Dennis Toddenroth, Jan Christoph, Ines Leb, Igor Engel, and Martin Sedlmaier, “A scoping review of cloud computing in healthcare”, Griebel et al. BMC Medical Informatics and Decision Making (2015).

[3] S. S. V. Chandra and C. G. Raji, “Artificial Neural Networks in Prediction of Patient Survival after Liver Transplantation”, vol. 7, no. 1, pp. 17, [2016].

[4] C. Hervas Martinez, and M. De La Mata, “Predicting Patient Survival after Liver Transplantation Using Evolutionary Multi Objective Artificial Neural Networks”; vol. 58, no. 1, pp. 3749 [2013].

[5] S. S. V. Chandra and C. G. Raji, “Predicting the survival of graft following liver transplantation using a nonlinear model”; Journal of Public Health October 2016.

[6] M. Vivareli and A. D. Pinna, “Artificial Neural Network is Superior to MELD in Predicting Mortality of Patients with End-Stage Liver Disease.”; Liver, www.gutinl.com [2007].

[7] Sanjay P. Ahuja, Sindhu Mani & Jesus Zambrano, “A Survey of the State of Cloud Computing in Healthcare”, Network and Communication Technologies; 2012.

[8] C. G. Raji and S. S. Vinod Chandra, “Various Medical Aspects of Liver Transplantation and its Survival Prediction using Machine Learning Techniques”, Indian Journal of Science and Technology, Vol 10(13), DOI: 10.17485/IJST/2017/v10i13/94111, April 2017.

[9] L. Bertocco de Paiva Haddad, L. Ducatti, L. R. B. Carelli Mendes and W. Andraus, and L. A. C. D’Albuquerque, “Predictors of micro-costing components in liver transplantation”, 2017 CLINICS DOI: 10.6061/clinics/2017.

[10] M. D. Ayllon, and R. Ciria, “Validation of Artificial Neural Networks as a Methodology for Donor-Recipient Matching for Liver Transplantation”, Liver Transplantation 24 192–203 2018 AASLD 2017.