Fuzzy Logic-Based Paddy Yield Prediction to Facilitate Weather Index-Based Crop Insurance

M. K. Dulanga Kithmini
Department of Industrial Management
University of Kelaniya
Sri Lanka
dulangakithmini@gmail.com

R. A. C. P. Rajapakse
Department of Industrial Management
University of Kelaniya
Sri Lanka
chathura@kln.ac.lk

***Abstract*—Weather risk can be considered as the dominant among the risks associated with agriculture. The unpredictability of weather and its adverse impact on agriculture has made the industry unstable. However, the authorities have not been able to provide farmers with a mechanism which assures that their livelihood is preserved. To address this issue, Weather index-based crop insurance has been used in many countries. Risk modeling is an important component in this insurance. After quantifying the uncertainties in crop harvest due to weather factors, the insurance company can guide the farmers accordingly. Crop yield prediction based on weather factors can be used for this. This study proposes a Fuzzy Logic System which relates input (weather factors) and output (crop yield) variables through linguistic rules. The output of this model can be used for risk modeling in weather index-based crop insurance based on which the possible paddy crop damage can be minimized.**

Keywords— Crop yield prediction, Fuzzy logic, Rule-based systems

I. INTRODUCTION

Farmers in developing countries are vulnerable to a range of risks. Weather risk can be considered as the dominant among those. The unpredictability of weather and its adverse impact on agriculture has made the industry an unstable sector since there is no assurance of return on the investment. Even though the weather risk has been prevalent throughout years in agriculture, the authorities have not been able to provide farmers with a mechanism which assures that their livelihood is preserved. The increasing severity of weather events associated with climate change constantly affects the livelihood of farmers. Banks are unlikely to lend to farmers as they think that the farmers cannot pay back the loans. The farmers are also unwilling and unable to invest money as it might affect their livelihood, if something goes wrong with the climate changes. This issue has impacted in a wide range of socio-economic aspects which ultimately has resulted in shaking the strongest pillar of the Sri Lankan economy, “Agriculture.”

A. *Paddy Cultivation*

While agriculture industry is concerned with a vast variety of crops, paddy cultivation is responsible for the most part of the industry given the fact that rice happens to be the staple food of more than half of the world’s population. More than 3.5 billion people depend on rice for more than 20% of their daily calorie requirement. Asia accounts for 90% of global rice consumption. Rice is the fastest growing food staple in Africa and is one of the fastest in Latin America. Global rice consumption remains strong, especially in many Asian and African countries [1].

Rice consumption in Asia exceeds 100 kg per capita annually in many countries. In terms of food consumption, Asia greatly depends on rice [2].

Per capita rice consumption has started to decline in high income Asian countries such as Republic of Korea and Japan. However, nearly a one fourth of the Asian population still has a considerable demand for rice [3].

Rice is the staple food in Sri Lanka and paddy cultivation is therefore given the utmost importance in the agriculture industry. Paddy is cultivated as a wetland crop in all the districts in Sri Lanka and paddy cultivation is the livelihood of about 1.8 million farm families island wide. Paddy crop is cultivated approximately in 870,000 Hectares at present which occupies 34 percent of the total cultivated area in Sri Lanka. Sri Lanka satisfies around 95 percent of the domestic requirement with the production of 2.7 million of rough rice. The per capita consumption of rice fluctuates around 100 kg per year depending on the price of rice, bread and wheat flour [4].

B. *Weather Index-based Crop Insurance*

For most of the properties, business entities and even lives which are associated with risks have insurance mechanisms to assess risks and provide coverages. Even though agricultural industry is also associated with higher levels of risks as mentioned above, there has not been a proper insurance mechanism to assess the crop damage and compensate accordingly due to the complexity of risk factors. To address this issue Weather index-based crop insurance has been used in many countries.

Weather index-based crop insurance is a tool to manage climate risk which enables to stabilize farmers’ income by providing timely payouts, considering weather factors. Weather index-based crop insurance is an approach to manage weather and climate risk because it uses a weather index, such as rainfall, to determine payouts. With weather index-based crop insurance contracts, an insurance company does not need to visit the policyholder to assess damages and arbitrate claims. Instead, if the rainfall recorded by gauges is below a previously agreed threshold, the insurance pays out automatically [5].

Faster payouts mean that the farmers do not have to sell their assets to survive. Weather index-based crop insurance is an alternative to traditional crop insurance which enables to support the farmers in developing countries [6].

Risk modeling is another important component in weather index-based crop insurance. This quantifies the uncertainties in crop harvest due to weather factors and is used to support decision making and planning in agriculture. After a proper risk assessment, the insurance company is guiding the farmers accordingly, before starting the cultivation process. Farmers are advised to cultivate less and actions are taken to minimize the risk of crop damage. This enables the farmers to have the maximum harvest with the resources available [7].

Risk modeling is the most challenging aspect in implementing the weather index-based crop insurance. In order to tackle this issue in an effective manner, crop yield prediction based on the weather factors can be used.

The purpose of this research is to facilitate risk modeling for implementing Weather index-based crop insurance for paddy farmers in Kurunegala District. A crop yield forecasting model for assessing risks and taking actions accordingly is needed. If the resources are not sufficient initially and the predicted crop yield is also low, the farmers can be guided to cultivate in a smaller area, without cultivating the whole area available. This enables the farmers to use the available resources efficiently and to reduce crop damage. The insurance companies do not have to give payouts if the crop damage can be controlled. Therefore, crop yield prediction enables the insurance companies to guide the farmers to use available resources efficiently and to take precautions to reduce crop damage.

Farmers, insurance companies which provide weather index-based crop insurance and the government will be benefitted from this research. Further, by this research the instability of the industry can be addressed, thus resolving a range of socio-economic issues such as deviation of the younger generation from the agricultural industry, poverty and food insecurity.

This research work can be carried out using statistical approaches, deep learning approaches or rule-based approaches. When using statistical approaches, complex and nonlinear relationships cannot be captured. When using deep learning approaches, a huge amount of data is required even though complex relationships and long-term dependencies can be captured. A deep learning approach is not suitable for this research since only 24 years of data is available. The model can be overfitted and the results may be biased when using small amount of data. When using fuzzy logic, it is able to handle linguistic concepts and is able to perform non-linear mapping between inputs and output. Also, it gives output considering different intermediate possibilities. Therefore, it is suitable in this research context.

Through this research work, a fuzzy logic-based paddy yield forecasting model is proposed to facilitate weather index-based crop insurance in Kurunegala district.

II. Literature Review

A crop yield is basically affected by the climate and finding the impact of environmental changes on crop yield has become important. Research works have been carried out on forecasting the crop yield using many different techniques. Following is a detailed review of literature on crop yield prediction in rule-based, deep learning and statistical approaches.

A. Rule-based Approach

Fuzzy logic is capable of providing an output considering all the intermediate possibilities. Therefore, it is appropriate for data which consist with uncertainties and nonlinearities. Many research works have been carried out using fuzzy logic due to its suitability for ambiguous expression of classification.

[8] A fuzzy logic model is proposed for the prediction of spontaneous combustion of coal in underground coal mines. CO, O_2, N_2 and temperature are used as the input variables and fire intensity is considered as the output variable. Fuzzy rules map the relationship between antecedents and consequent. In this study, “MIN” operator is used to combine the fuzzy relations. The model is simulated by using the Mamdani inference system and is run by the Fuzzy Logic Toolbox in MATLAB.

The results show reliable predictions and the fuzzy logic system has been able to capture the uncertainties and nonlinearities of the data. The study has been able to identify gas station points which have a higher chance of causing spontaneous combustion.

[9] A fuzzy inference system has been proposed for sorghum yield prediction. Physio-morphological features such as plant height, panicle length, panicle weight etc. have been considered for developing the model. Different combinations of inputs and outputs have been used and prediction accuracy is measured by Root Mean Square (RMS) value. One-to-one mappings for all the input parameters and combinations of parameters have been considered. Linguistic variables have been given to each parameter and triangular membership functions have been used. The rule base consists of ten rules and pairs of inputs are considered based on their mutual inter-relationships.

By using the fuzzy inference system, accurate results have been given with an RMS ranged between 1.39 and 3.9, even though the rule base consisted of only 10 rules. The predicted values have been matched with the actual values with low variation.

[10] An Adaptive Neuro-Fuzzy Inference System (ANFIS) is proposed for crop yield forecasting using Sugeno-Kang approach. The model is developed using climatic variables (temperature, rainfall, evaporation, humidity etc.) and crop yield. Triangular membership functions have been adopted and the learning capability of an artificial neural network has been used for automatic fuzzy rule generation. MATLAB is used to build the model and the model is validated by coefficient of correlation.

In this research work, more refined fuzzy rules have been created since the rules are generated through Artificial Neural Networks.

B. Deep learning Approach

Artificial Neural Networks (ANN) are capable of capturing the patterns in data by auto-tuning weights through back propagation. This has enabled to use Artificial Neural Networks for capturing dynamic and nonlinear data patterns.

Long Short-Term Memory (LSTM) is a special type of Recurrent Neural Networks (RNN) which are specialized in capturing long-term dependencies. Therefore, research works have been carried out using these technologies when nonlinearities and long-term dependencies in data are involved.

[11] A Long Short-Term Memory (LSTM) based aggregated model has been developed for air pollution forecasting. The basic idea of this model is similar to that of voting. Three input layers are provided with local data sets such as local emissions of pollution and air quality, near station data set and with industrial stations data sets. The model has been evaluated by comparing with SVR (Support Vector Machine based Regression), GBTR (Gradient Boosted Tree Regression) and LSTM using various assessment techniques such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The aggregated model has been able to effectively improve the accuracy of prediction.

LSTM is a special type of Recurrent Neural Networks which is specialized in capturing long term dependencies. Therefore, an LSTM model would be suitable for time series forecasting. Independent variables used in this research do not guarantee to have a linear relationship with the output. LSTM model has been able to capture non-linear relationships as well. However, deep learning models require a huge amount of data. Therefore, the results can be overfitted and biased if a huge amount of data is not available.

C. *Statistical Approach*

[12] This research work produces three independent models to predict winter wheat yield. Multiple Linear Regression (MLR) method has been used to build the models. Meteorological data, information on mineral fertilization and yield data are used. Quality of the models has been determined using Relative Absolute Error (RAE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percent Error (MAPE).

According to the results of the research work, it has been noted that the yield prediction error increases with the increase of number of independent variables. Also, the models may not be able to capture the complex relationships between weather data and the crop yield.

Even though past studies have been carried out by using these approaches, statistical approaches and deep learning approaches are not suitable in this research context. Statistical models may not be capable in capturing complex relationships between weather factors and crop yield. A huge amount of data is required to build a deep learning model and the available data for this research is not sufficient for that. According to the literature, fuzzy logic-based models have been able to give reliable predictions even with a small set of fuzzy rules. Since fuzzy logic is capable of handling non-linearity and giving outputs considering different intermediate possibilities, a fuzzy logic-based system is appropriate for crop yield prediction.

III. Data Collection

Weather data (rainfall, temperature, humidity, sunshine hours and wind speed) of Kurunegala district were collected from the Department of Meteorology, Sri Lanka and paddy yield data was collected from the Department of Census and Statistics, Sri Lanka. 24 years of data from 1996 to 2019 were used for the research.

IV. Methodology

When predicting crop yield, it is better to give an output which consider all the intermediate possibilities. Fuzzy logic is appropriate in this context. Fuzzy theory is a mathematical theory which is related with uncertainty. This is suitable when ambiguity lies in the context. Fuzzy logic is based on fuzzy set theory, fuzzy rules and fuzzy reasoning. Fuzzy logic is the concept of fuzzy sets incorporated with degrees of truth. A fuzzy logic model consists of a fuzzifier, rules base, inference engine and a defuzzifier [13]. Fig. 1 shows the structure of a fuzzy logic system.

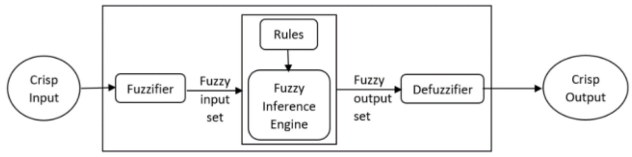

Fig. 1. Structure of a Fuzzy Logic Model

Fuzzifier is used to convert crisp input sets into fuzzy input sets. In this step, linguistic variables were given to antecedents and consequents based on the membership functions [8]. Fuzzy rules were extracted from the collected dataset and domain knowledge. Various operators can be used to combine the fuzzy relations. MIN operator was used in this research. Decision-making based on the rules is happening in the fuzzy inference engine. Fuzzy inputs are processed by using fuzzy logic theories and creates fuzzy outputs. In the defuzzification process, output fuzzy set is converted to a crisp output. Centroid of area was used for defuzzification.

Five weather factors (rainfall, temperature, humidity, sunshine hours and wind speed) and the sowing stage were considered for this research. ‘SciPy’ python package was used to find the pairwise correlation among all the columns using Spearman method. Features with low correlation were eliminated.

Prediction of paddy crop yield before the beginning of the paddy season is the main purpose of the model. Data from 1996 to 2017 were used for building the model and the data of 4 paddy seasons from 2018 and 2019 were used to test the model.

Temperature and sunshine hours in the sowing stage were eliminated in the feature selection process and rainfall, humidity and wind speed were considered in building the model. ‘skfuzzy’ fuzzy library was used.

Two models were built with 3 fuzzy levels (Low, Average and High) and 5 fuzzy levels (Very low, Low, Average, High and Very high). Fuzzy levels were defined using ‘automf’ function in ‘skfuzzy’ fuzzy library.

A. *Model 1*

A model was built with 3 fuzzy levels as {poor, average and good} for the input variables and {Low, Average and High} for crop yield per Hectare. Data range was selected between the recorded minimum value and the maximum value for creating the membership functions. Triangular membership functions were created. Fig. 2 and Fig. 3 show the membership functions of rainfall of the sowing stage and crop yield per Hectare respectively.

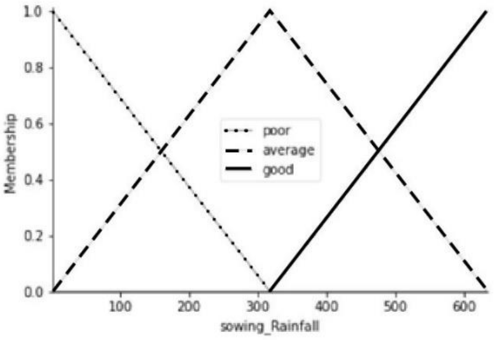

Fig. 2. Membership functions of rainfall in sowing phase – Model 1

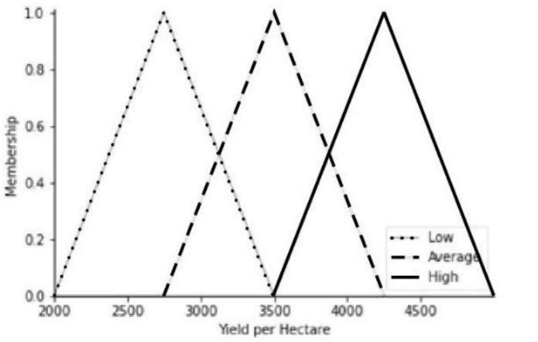

Fig. 3. Membership functions of yield per Hectare – Model 1

Fuzzy rules were created from the available dataset in the format of IF-THEN. Fig. 4 shows few rules from the rule base.

```
rule1 = ctrl.Rule ( sowing_Rainfall['average']  & sowing_Humidity['average'] & sowing_WindSpeed['poor'] , yield_Per_Hectare['Average'])
rule2 = ctrl.Rule ( sowing_Rainfall['average']  & sowing_Humidity['good'] & sowing_WindSpeed['average'] , yield_Per_Hectare['High'])
rule3 = ctrl.Rule ( sowing_Rainfall['average']  & sowing_Humidity['average'] & sowing_WindSpeed['average'] , yield_Per_Hectare['Average'])
rule4 = ctrl.Rule ( sowing_Rainfall['average']  & sowing_Humidity['good'] & sowing_WindSpeed['average'] , yield_Per_Hectare['High'])
```

Fig. 4. Fuzzy rules - Model 1

B. *Model 2*

A model was built with 5 fuzzy levels as {poor, mediocre, average, decent and good} for the input variables and {Very low, Low, Average, High and Very high} for crop yield per Hectare. Triangular membership functions and trapezoidal membership functions were created. Fig. 5 and Fig. 6 show the membership functions of rainfall of the sowing stage and crop yield per Hectare respectively.

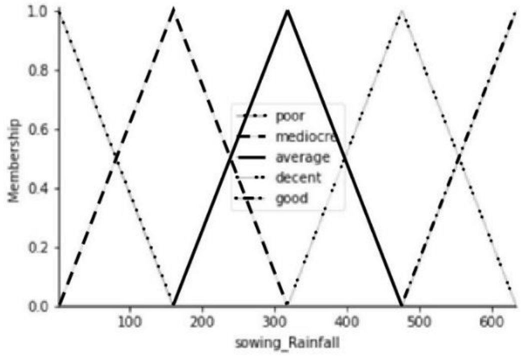

Fig. 5. Membership functions of rainfall in sowing phase – Model 2

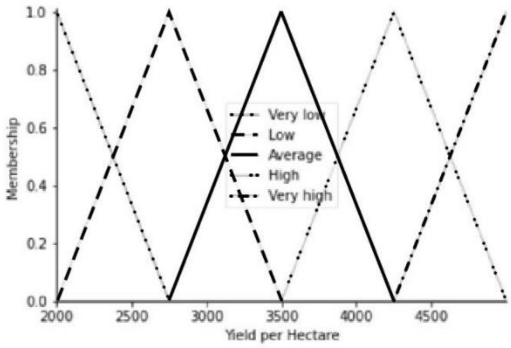

Fig. 6. Membership functions of yield per Hectare – Model 2

V. RESULTS AND DISCUSSION

Four paddy seasons of 2018 and 2019 were used to test the model. Fig. 7, 8, 9, 10, 11, 12, 13, 14 show the predicted yield from Model 1 and Model 2 in those paddy seasons.

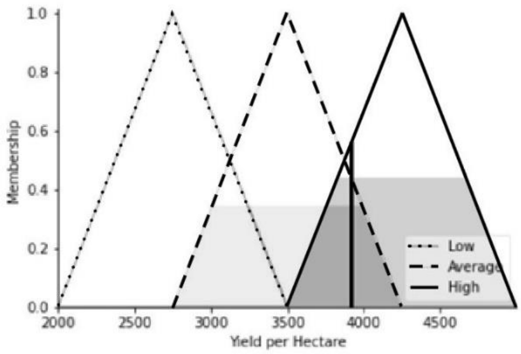

Fig. 7. Predicted yield of 2018 'Yala' season - Model 1

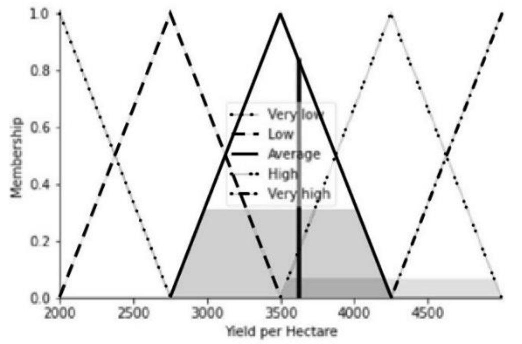

Fig. 8. Predicted yield of 2018 'Yala' season - Model 2

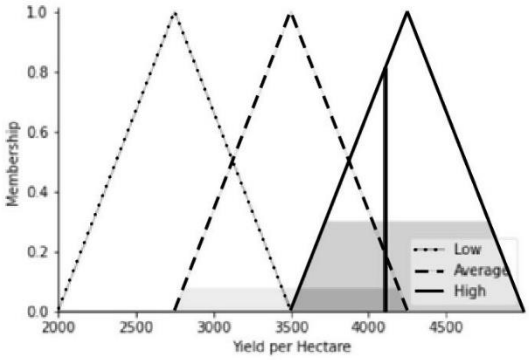

Fig. 9. Predicted yield of 2018 'Maha' season - Model 1

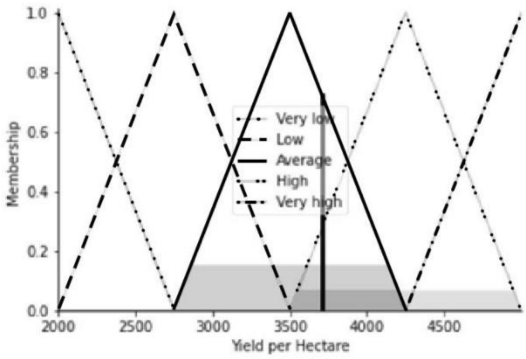

Fig. 10. Predicted yield of 2018 'Maha' season - Model 2

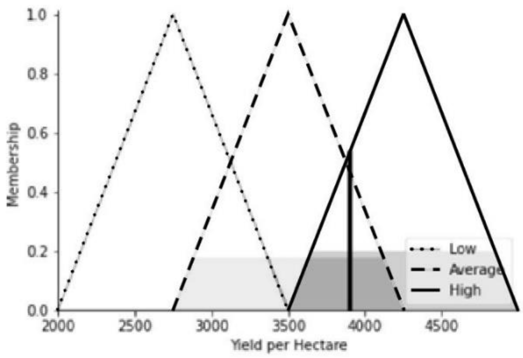

Fig. 11. Predicted yield of 2019 'Yala' season - Model 1

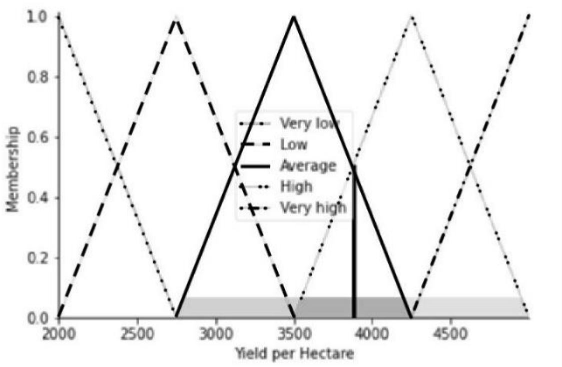

Fig. 12. Predicted yield of 2019 'Yala' season - Model 2

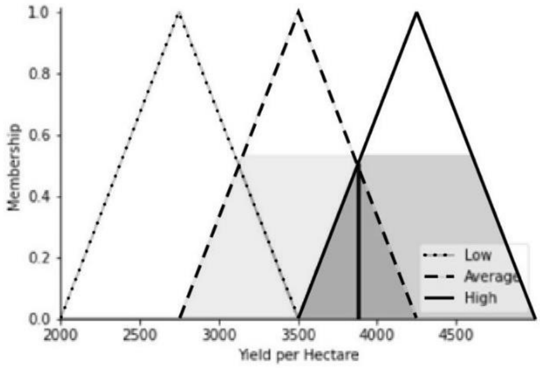

Fig. 13. Predicted yield of 2019 'Maha' season - Model 1

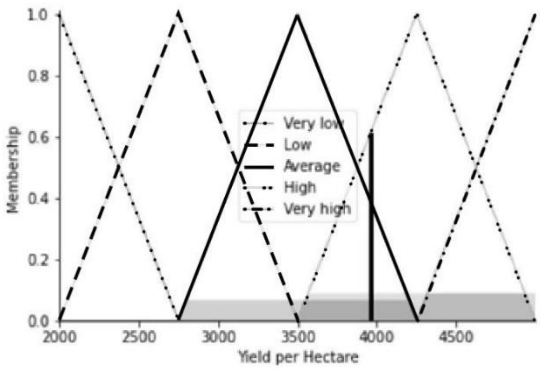

Fig. 14. Predicted yield of 2019 'Maha' season - Model 2

Fig. 15 and Fig. 16 show the representation of actual yield and predicted yield from Model 1 and Model 2 respectively.

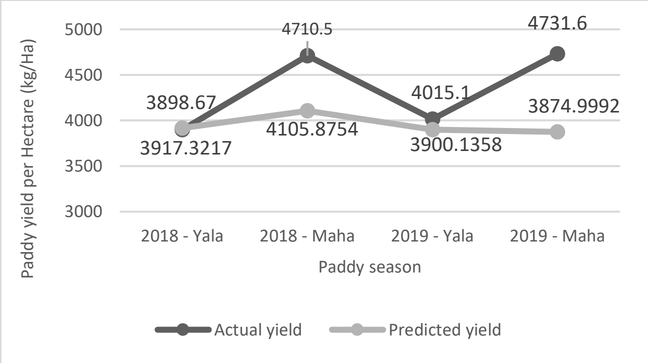

Fig. 15. Graphical representation of actual yield and predicted yield - Model 1

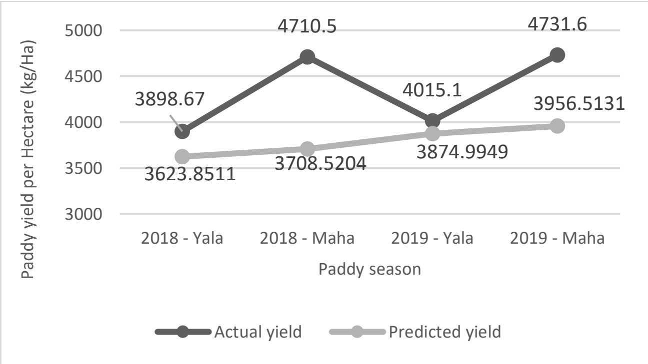

Fig. 16. Graphical representation of actual yield and predicted yield - Model 2

Graphs in Fig. 15 and Fig. 16 show the preliminary results of the models. Model 1 has given more reliable results than Model 2 and there is a slight variation in the predicted paddy yield from the actual paddy yield. In terms of predicted fuzzy levels, Model 1 has been given more reliable results.

Model 1 (with 3 fuzzy levels) predicted with a Mean Absolute Error of 399 kg/Ha and Model 2 (with 5 fuzzy levels) predicted with a Mean Absolute Error of 548 kg/Ha. Getting predictions with a Mean Absolute Error of 399 kg/Ha can be considered accurate since the results are used to get predictions before even starting the paddy season.

There are various intermediate possibilities related to weather factors and crop yield. Therefore, implementing a fuzzy logic system is suitable for this context. Mean Absolute Error values recorded prove that the model is capable of accurate prediction and is suitable for risk modeling to facilitate weather index-based crop insurance.

VI. Conclusion

In this study, a Fuzzy Logic System which can predict the paddy crop yield is proposed. Data were collected from the Department of Meteorology, Sri Lanka and the Department of Census and Statistics, Sri Lanka. Rainfall, humidity and wind speed of the sowing stage were the inputs and crop yield per Hectare with the relevant fuzzy level was the output. 'skfuzzy' fuzzy library by 'sklearn' was used to build the model. Two models were built with 3 fuzzy levels and 5 fuzzy levels and predicted output was compared with the actual output. Both models have been given accurate results while the model with 3 fuzzy levels has been able to output the correct fuzzy level as well. Finally, Mean Absolute Error was used to find the average error of the predicted values and it proved the reliability of the models. Existing models have used aggregated data of the whole crop season, disregarding the effect of weather conditions in the sowing stage. This model is more practical since this gives results considering the weather conditions in sowing stage.

VII. Future Work

This research was carried out considering the impact of weather factors on paddy yield. There are other factors such as fertilizers, diseases and soil conditions. More focus will be given on incorporating multiple factors that affect crop yield in the future work.

References

[1] Ricepedia, http://ricepedia.org/rice-as-food/the-global-staple-rice-consumers

[2] Ricepedia, http://ricepedia.org/rice-around-the-world/asia

[3] Food and Agriculture Organization of the United States, http://www.fao.org/docrep/003/x6905e/x6905e04.htm

[4] Department of Agriculture, https://doa.gov.lk/rrdi/index.php?option=com_sppagebuilder&view=page&id=42&lang=en

[5] J.D. Daron, D. S. Assessing pricing assumptions for weather index insurance. Climate Risk Management, 76-91 (2014)

[6] Climate Change, Agriculture and Food Security, https://ccafs.cgiar.org/fr/node/50640#.XyXOZCgzZPZ

[7] Stoppa, W. D., Jamie Anderson, E. C., & Rispoli, F. Weather (Raí A. Schwalbert, Geomar Corassa, & P.V.Vara Prasad, 2020) index-based Insurance in Agriculture Development. International Fund for Agricultural Development (2011)

[8] Esmatullah Danish, M. O. Application of Fuzzy Logic for Predicting of Mine Fire in Underground Coal Mine. Safety and Health at Work (2020)

[9] M. A. Jayaram, N. M. Fuzzy Inference Systems for Crop Yield Prediction. Journal of Intelligent Systems (2012)

[10] Kalpesh Borse, P. G. Prediction of Crop Yields Based on Fuzzy Rule-Based System (FRBS) Using the Takagi Sugeno-Kang Approach. In I. Z. Pandian Vasant, & G.- W. Weber, Intelligent Computing & Optimization (pp. 438-447). Springer (2019)

[11] Yue-Shan Chang, H.-T. C., Satheesh Abimannan, Y.-P. H., & Yi-Ting Tsai, K.-M. L. An LSTM-based aggregated model for air pollution forecasting. Atmospheric Pollution Research, 1451-1463 (2020)

[12] Niedbala, G. Application of multiple linear regression for multi-criteria yield prediction of winter wheat. Research and Applications in Agricultural Engineering (2018)

[13] Toshiro Terano, K. A., & Sugeno, M. Fuzzy Systems Theory and its Applications. London: Academic Press Limited (1992)