|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **COMSATS University Islamabad** | | | |
|  |  | **Campus** |  |
| **Registrar Office, Principal Seat, Islamabad** | | | |

Form for Approval of Synopsis of Graduate Programs Students (MS/MBA/Ph.D.) of CUI

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ***Student’s Details:*** | | |  | |  |  | |
| **Registration No:** | | SP22-RCS-003 | | | **Name:** | Abdul Hadee Anwaar | |
| **Program Name:** | **MS(Computer Science)** | | | | **Father Name:** | | Muhammad Asif Askari |
| **Area of Specialization:**  (if any as per approved SoS) | | | |  | | | |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ***Supervisor/ Co-Supervisor’s Details:*** | | |  | |  |  |
| **Supervisor’s Name & Designation:** | | Dr- Faisal Azam | | | | |
| **Co-Supervisor’s Name & Designation:** | | | |  | | |
| **Synopsis Title:** |  | | | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| ***Recommended & Signed by Supervisory Committee: (Name & Signature)*** | | |  |
| 1. |  | Member/Supervisor | |
| 2. |  | Member/Co-Supervisor | |
| 3. |  | Member | |
| 4. |  | Member | |

|  |  |  |
| --- | --- | --- |
| ***Recommended & Signed by Departmental Graduate Advisory Committee: (Name & Signature)*** | |  |
| 1. |  | |
| 2. |  | |
| 3. |  | |
| 4. |  | |
| 5. |  | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Signed by Student:** |  | | |
| **Signed & Recommended by HoD**  (on the basis of Turnitin Similarity Report (attached): | |  | |
| **Signed and approved by the respective Dean:** | | |  |

|  |  |
| --- | --- |
|  | **COMSATS University Islamabad** |
| **\_\_\_\_\_\_\_\_\_\_\_\_ Campus** |
| **Registrar Office, Principal Seat, Islamabad** |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Synopsis** for the degree of |  |  | M.S./M.Phil |  | Ph.D. |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***PART-1*** *(to be completed by the student):* | | | | | | |  |  | |
| **Registration No:** | | SP22-RCS-003 | | | | | **Name:** | Abdul Hadee Anwaar | |
| **Department:** | **ComputerScience** | | | | | **Date of Thesis Registration:** | | |  |
| **Title of Research Proposal:** | | | | |  | | | | |
| **Research Supervisor:** | | | Dr Faisal Azam | | | | | | |
| **Research Co-Supervisor:** | | | |  | | | | | |

|  |  |  |  |
| --- | --- | --- | --- |
| ***Members of Supervisory Committee:*** | |  | |
| 1. |  | | |
| 2. |  | | |
| 3. |  | | |
| 4. |  | | |
|  |  | | |
|  | *Student’s Signature* | |  |
|  |  | |  |

**Summary of the Research**

The main objective of this study is to develop a high-accuracy flood prediction model using machine learning techniques. Flooding is a significant natural disaster that causes significant damage to properties, infrastructure, and human lives. The scale of this catastrophic event can be predicted calculated by observing the overall its impacts.

Only in Pakistan alone flood claimed more than 1700 lives and effected 33million people just in the year 2022(Assessment, 2022). India has also been flood victim for past few years with vast destruction in last few years(*RS\_Session\_246\_AU\_2225*, n.d.). Accurate prediction of floods is critical to reduce the risks and consequences of flooding. Many models ANN, MARS ,LSSVR, SVRRF, DT and many others are working throughout the world with promising accuracies(Technology, 2022) but are under specific conditions. Therefore, developing an effective flood prediction model is essential.

The developed model will be evaluated using metrics such as mean squared error, root mean squared error, and R-squared. These metrics will provide an accurate assessment of the model's performance and help determine its accuracy. The evaluation process will ensure that the developed model is effective in predicting floods accurately.

The study also aims to compare the developed model's performance with other machine learning models for flood prediction. Comparing the developed model with other models will provide valuable insights into the most effective machine learning techniques for predicting floods. This comparison will also highlight the strengths and weaknesses of the developed model and identify areas for improvement.

Additionally, this study's objective is to determine the most significant attributes that affect flood prediction using correlation matrix analysis. After visualizing and understanding the most significant attributes that affect flood prediction is essential for developing an effective flood prediction model. By identifying the most significant attributes, the developed model will be able to predict floods more accurately.

Moreover, this study intent to contribute to the field of flood prediction by providing valuable insights for researchers, authorities, and individuals. By developing an accurate flood prediction model, this study will provide a useful tool for mitigating the risks and consequences of flooding. The study's findings will also contribute to the existing knowledge of flood prediction and provide a basis for future research in this field.

In summary, this study aims to develop an accurate flood prediction model using machine learning techniques, evaluate its performance using metrics, compare it with other models, determine the most significant attributes that affect flood prediction, and contribute to the field of flood prediction.

1. **Introduction**

Natural disasters like floods have the potential to be fatal and cause significant damage to things like buildings and highways. Over the past 40 years, floods in the western United States have cost more than $1 billion in damages annually on average. Climate change, rapid city growth, and forest destruction are all contributing to an increase in the frequency and severity of floods. It's crucial to create precise models that can forecast when and where floods may occur. In this approach, governments and the general public can receive early warnings and take action to protect themselves and minimize harm. Predicting floods correctly can help make them less destructive and hazardous.

This study attempts to evaluate the precision of machine learning-based flood predictions. We want to specifically forecast the height of the flood levels using a linear perceptron model, and then evaluate the quality of the model using metrics like mean squared error (MSE), root mean squared error (RMSE), and R-squared (R2). To determine how effective the linear perceptron model is, we will compare its performance to that of other machine learning models.

Our research aims to create precise flood prediction models that can aid in lowering flood risks and their effects. With the help of these models, local governments in flood-prone areas can better allocate their resources and residents can take the necessary safety measures. These models can also help policymakers by educating them on the best flood control methods.

The complexity and unpredictability of weather patterns, topography, and water movement make it challenging to develop an efficient model, despite numerous attempts to do so. However, improvements in technology and data analysis have made it feasible to make flood predictions that are more precise. Our objective is to develop a model utilizing machine learning methods that can offer accurate flood forecasts in real-time, enabling government agencies and citizens to act immediately and reduce the risks and effects of floods.

Our methodical approach will include a number of processes, including data cleaning, visualization, correlation matrix analysis, test-train split, and linear perceptron modelling. To do this, we will make use of the Flood Prediction datasets, which are made up of historical flood data from the Chinese region around Poyang Lake. This dataset contains a number of variables, including rainfall, water level, and water flow, that can be used to forecast floods.

By creating a precise and useful flood prediction model, our study intends to advance the field of flood forecasting. Authorities and people can use the study's findings to plan for floods and lessen their effects. Additionally, this study can offer helpful insights for academics working in the fields of data analysis and machine learning.

The study's goal is to use machine learning to develop an accurate and reliable flood prediction model in order to meet the growing worry about floods. By correctly anticipating floods, we can lessen their dangers and effects while also assisting people and government agencies in taking prompt action to prepare and lessen the effects of floods. Our research has the potential to advance the science of flood forecasting and offer insightful information to individuals, organizations, and scholars.

1. **Research Objectives**

* Create a very accurate flood prediction model using machine learning.

Utilize metrics such as mean squared error, root mean squared error, and R-squared to assess the effectiveness of the generated model.

* Compare the effectiveness of the created model with other flood prediction machine learning models.
* Utilize correlation matrix analysis to identify the most important characteristics that influence flood forecast.
* By giving academics, authorities, and people useful information, you can make a contribution to the field of flood forecasting.

These simple and concise research objectives highlight the study's major objectives. To achieve high accuracy, the proposed machine learning-based flood prediction model will be assessed using a variety of criteria. This study intends to offer insights into the most effective approaches for flood prediction by comparing the generated model's performance with other models and examining the most important aspects. Additionally, through offering important information to researchers, authorities, and people, this study hopes to advance the field of flood prediction.

1. **Related Work**

Due to eminence impact of floods, much work has been performed for its prediction and forecasting previously. Literature provides us with extensive and in-depth work on this topic to get an idea of the already performed work on it.

Flood relevant data can be seen in various form as its information can be gathered from various resources. (Puttinaovarat, 2020) used big data obtained from various resources to be used. Extracting flood relevant information from images is also a popular approach adopted by various researchers(Lima & Pereira, 2023) (Ouma & Omai, 2023) (Mastro et al., 2022). Numerical data has been under the light as it can store huge amount of data in a manageable way especially in case of climate data, numerical way is considered a systematic way. It has been used frequently in current times to achieve the research goals(Cai et al., 2022)(Baguis et al., 2022)(Declercq et al., 2023). The most important means of information whose importance can’t be neglected in any field apart from flood forecasting is the use of social media information. Social media is the fastest way of attaining authentic and valid data relevant to a particular topic(Kanth et al., 2022). The extensive use of social media for data attaining has been very popular for quite some time now especially when we put light for flood relevant fields(Khan et al., 2022)(Ouyang et al., 2022)(Ouyang et al., 2022)(Kanth et al., 2022).

The type of input data is selected while keeping in view the models that will be fed with that data. Researchers have used variety of models to attain the most accurate results for flood forecasting. Due to the advancements and progress in the research area now we can apply a variety of models to achieve our goal. Talking at a bigger level various Machine Learning (ML) and Deep Learning (DL) models have assisted in the past for research purposes. To forecast floods, artificial neural networks (ANNs) are trained on historical flood data and environmental factors including rainfall, water level, and river flow(Rifat & Liu, 2022). SVMs predict floods based on weather and hydrological data and can analyze complex nonlinear relationships between variables(Zehra, 2020). Long Short-Term Memory (LSTM) models, a type of recurrent neural network, have been successfully used in flood prediction research for time series forecasting of river flow data. LSTMs are known for their ability to handle long-term dependencies in the data and have shown promising results in the field of flood prediction(Ghimire et al., 2021). Capsule Networks have been utilized in recent research for rainfall-runoff modeling and flood prediction. These networks have shown to be useful for handling variable-length input data and detecting patterns in the data. Capsule networks have the potential to improve the accuracy of flood prediction models by capturing the spatial relationships between features and can be further explored in future research for improving flood prediction systems(Runo et al., 2020).

Combination of different models have been seen to give improved and in-depth results as compared to stand alone models. This practice is very popular as many researchers have opted for these. (Zannou et al., 2021) used a hybrid of the ant colony optimisation (ACO) method and artificial neural networks (ANNs) to anticipate floods. Their strategy includes combining past flood data and meteorological data to train the model. Artificial neural networks (ANNs), genetic algorithms (GA), and particle swarm optimisation (PSO) have all been proposed as components of a hybrid artificial intelligence system for flood prediction. The model was trained and assessed using historical flood data as well as meteorological data(Zannou et al., 2021). (Abbaszadeh Shahri et al., 2022) used deep neural networks (DNNs) to predict floods and added uncertainty quantification to boost the precision of their forecasts. Both historical flood data and meteorological data were used to train the model. (Sun et al., 2022) provides a strategy for forecasting floods using a multi-model ensemble. The study makes use of five different models and historical hydrological data from the Chinese Dong ting Lake basin. The results demonstrate the importance of including uncertainty in flood predictions and demonstrate that the method is more accurate and robust than individual models. (Chen et al., 2022) this piece of research mixes artificial neural networks (ANNs) and adaptive neuro-fuzzy inference systems (ANFIS) to suggest a hybrid intelligent system for flood forecasting. Utilizing historical hydrological data from the Indonesian Tabak River Basin, the system was trained and tested, and the findings show increased accuracy in flood prediction. In areas with a lack of data, the study emphasizes the potential of combining various artificial intelligence techniques for effective flood forecasting.

[1000 to 1500 words approx.]

1. **Problem Statement**

Our work tackles the urgent problem of flood prediction, which has gotten worse recently as a result of the increased severity and frequency of floods. Our research focuses on the shortcomings of current flood prediction models and the additional elements that have evolved that reduce the efficacy of these models. The problem is made worse by out-of-date equipment and a dearth of trustworthy data. We want to address these issues by developing a cutting-edge flood prediction model that can offer precise and timely predictions, lessening the impact of floods. Our objective is to increase the accuracy of flood predictions and lessen the harm that floods do.

1. **Research Questions (if any)**
2. ow can geographic information be added to flood prediction models to increase their accuracy and dependability?
3. How can outmoded technology be changed or improved to improve the ability to predict floods?
4. What effect does real-time data have on the precision of flood predictions, and how can this data be processed efficiently to enhance predictions?
5. What is the best procedure for creating a flood prediction model that can handle massive quantities of data in real-time and scale?
6. How can the created flood prediction model be verified and assessed to guarantee its veracity and dependability?

This study's key research question is: How can a unique flood prediction model be created that efficiently uses region-specific data, makes use of real-time data processing, and enhances the accuracy and timeliness of flood predictions?licit research question(s) remove this section from the synopsis.

1. **Research Methodology**

The proposed research methodology will involve a quantitative approach to address the problem statement and research objectives. The following steps will be followed:

* Data Collection: Keeping in view our final goal we have attained a dataset that have many attributes that can affect or play the part in flooding. We have used extensive data as short term data can become biased.

* Data Pre-processing: The collected data was not clean. We performed many pre-processing’s steps like removal of null values, duplicated values were eliminated, Rounding of the values. This provided us with a clean data, that was used for further work.
* Data Analysis: Once clean data was obtained then we started to analyze data and diagnose different relationships between the attributes and our final objective, Correlation Matrix, histograms, scatter plot and box plots were under the observation for extracting any useful relation for our cause.

* Model Selection: After analyzing the data, different machine learning algorithms such as linear regression, decision trees, random forests, and neural networks were under the observation but we opted for linear perceptron as it was not widely used before and was considered due to its problem solving efficiency.

* Model Training: The selected machine learning model was trained on the preprocessed data. We split the preprocessed data to train the linear perceptron model then the leftover values were used for testing.
* Model Evaluation: The trained model will be evaluated on a separate test dataset to measure its accuracy and performance.

* Performance Evaluation Measures: The performance of our model is evaluated using metrics such as accuracy, precision, recall, and F1 score as they are considered the best parameters for identification of a model performance.

* Result Analysis and Visualization: The final step is to analyze the results and visualize the predictions. The results will be compared with existing models and evaluated for their effectiveness in flood prediction.

To carry out the research, a large dataset containing various attributes such as rainfall, water level, and temperature was used. The performance of the model will be evaluated on different evaluation measures, and the results will be analyzed and visualized using various data visualization techniques.

1. **References**

(Use IEEE standard referencing style)

<http://www.ieee.org/documents/ieeecitationref.pdf>

1. **Tentative Schedule**

(Identify major milestones and deliverables along with schedule)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tasks** | **Date/ Duration** | **Date/ Duration** | **Date/ Duration** | **Date/ Duration** | **Date/ Duration** | **Date/ Duration** |
| Task-I |  |  |  |  |  |  |
| Task-II |  |  |  |  |  |  |
| Task-III |  |  |  |  |  |  |
| Task-IV |  |  |  |  |  |  |
| Task-V |  |  |  |  |  |  |
| Task-VI |  |  |  |  |  |  |

**PART II**

Recommendation by the Research Supervisor

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Name:** |  | **Signature:** |  | **Date:** |  |

**Signed by Supervisory Committee**

|  |  |  |  |
| --- | --- | --- | --- |
| **S. #** | **Name of Committee Member** | **Designation** | **Signature & Date** |
| 1 |  |  |  |
| 2 |  |  |  |
| 3 |  |  |  |
| 4 |  |  |  |

**Approved by Departmental Advisory Committee**

Certified that the synopsis has been seen by members of DAC and considered it suitable for putting up to BASAR.

Secretary

Departmental Advisory Committee

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Name:** |  | **Signature:** |  | **Date:** |  |

Chairman/HoD:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Name:** |  | **Signature:** |  | **Date:** |  |

**PART III**

**Dean, Faculty of Information Sciences & Technology**

|  |  |
| --- | --- |
|  | Approved for placement before BASAR. |
|  | Not Approved on the basis of following reasons |
|  |  |
|  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Signature:** |  |  | **Date:** |  |

**Secretary BASAR**

|  |  |
| --- | --- |
|  | Approved from BASAR. |
|  | Not Approved on the basis of following reasons |
|  |  |
|  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Signature:** |  |  | **Date:** |  |

**Dean, Faculty of Information Sciences & Technology**

|  |
| --- |
|  |
|  |
|  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Signature:** |  |  | **Date:** |  |

**List of courses studied**

(This must be provided on a separate Page)

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr. #** | **Course Code** | **Course Title** | **Semester** |
| **1** | CSC510 | Theory of Computation | FA21 |
| **2** |  |  |  |
| **3** |  |  |  |
| **4** |  |  |  |
| **5** |  |  |  |
| **6** |  |  |  |
| **7** |  |  |  |
| **8** |  |  |  |
| **9** |  |  |  |

**Guidelines for Synopsis Writing**

* The synopsis should follow scientific writing principles. For help and guidance read: [Science research writing for non-native speakers of English](https://www.pdfdrive.com/science-research-writing-for-non-native-speakers-of-english-e165857168.html).
* The whole synopsis must carefully be proofread for grammatical mistakes and typos. Students must submit signed synopsis completed in all aspects. In addition, a plagiarism report must be attached at the end of the synopsis.
* The Title of the synopsis should be selected carefully because it cannot be changed later, and it must reflect the objectives of the study. It should be focused and oriented to exhibit the nature of the study.
* For Heading Level-1, use Font: Times New Roman, Size 12, Bold.
* For Heading Level-2, use Font: Times New Roman, Size 11, Bold.
* For Heading Level-3, use Font: Times New Roman, Size 11, Bold, Italicize.
* All Headings/Sub-headings must be properly numbered except “Summary of the Research”.
* Text in each section must follow Font: Times New Roman, Size 12 with Line Spacing:1.5.
* All Figures and Tables must be properly captioned in camel notation. For Figures, the caption must be at the bottom whereas, for Tables, the caption must be at the top. The caption of the Figure/Table must be of Font: Times New Roman, Size 10, See Example Below:

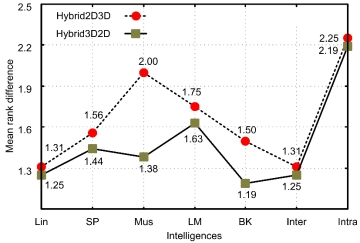


Figure. 1. Intelligences Mean Rank Differences

Table 1.  List of Methodologies used for Classifying WBC datasets

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Author | Year | Methods | Method Type | Dataset | Accuracy |
| Kutlu et al [16] | 2020 | Resnet50-CNN model with selective search, SSD, YOLOV3 deep models | Detection + Classification | BCCD, LISC | 97.52% |
| Baink et al [37] | 2020 | Color space transformation, K-mean clustering algorithm, CNN | Segmentation + Classification | BCCD | 98.61%, 96.00% |
| Gupta et al [32] | 2019 | OBBA, KNN, Logistic regression, RF, Decision based Tree classification | Classification | LISC | 97.30% |

* All the Figures and Tables must be properly cross-referenced and explained in the text.
* The content inside tables must be of Font: Times New Roman, Size 10.
* The figures must have good resolution and the content of the figure must be readable.
* There must be some text between two consecutive headings/ sub-headings.
* For References use IEEE standard referencing style and citation must be accordingly. For reference use Font: Times New Roman, Size 11 with Line Spacing: Single
* Carefully read all the items in the front pages and last pages and fill them accordingly.
* Before submission the synopsis must be duly signed by student, members of the supervisory committee, and members of the advisory committee.
* In the case of PhD synopsis, the word limit for each section can be slightly increased with the consent of the respective supervisor.
* Please specify the list of courses studied during the degree in the designated area in the synopsis.
* The literature review section should include research articles published in impact factor Journals or conferences with high impact score (https://research.com/conference-rankings/computer-science).