NATURALISTA CAMPANO
ISSN: 1827-7160
Volume 28 Issue 1, 2024

Developed a Smooth Support Vector Machine to Predict the Crop Production in Alluvial Soil and Red Soil Regions of Tamil Nadu, India

Dr. R. Tamilselvi, Dr. K. S. Mohanasathiya, Dr. S. Prasath
[1]*Assistant Professor & Head, School of Computer Science, VET Institute of Arts and Science (Co-Edu) College, Erode, Tamil Nadu, India.*
[2,3]*Assistant Professor, School of Computer Science, VET Institute of Arts and Science (Co-Edu) College, Erode, Tamil Nadu, India.*
Email: [1]*tamilselvircsc@vetias.ac.in*, [2]*Sathyaanandh08@gmail.com*, [3]*softprasaths@gmail.com*

Abstract: Crop production in Tamil Nadu, India, is influenced by various factors including soil type, climate, and agricultural practices. This study presents a comparative analysis of crop production in two prevalent soil types found in Tamil Nadu: alluvial soil and red soil. Alluvial soil, primarily located in river delta regions such as the Cauvery and Vaigai deltas, is known for its high fertility and good drainage properties, allowing it to be utilized for a variety of crops including rice, sugarcane, cotton, and various pulses and vegetables. On the other hand, red soil, predominant in central and western regions of Tamil Nadu, is less fertile but supports crops like millets, pulses, oilseeds, and certain fruits such as mangoes and guavas. Farmers in Tamil Nadu employ mixed cropping and crop rotation techniques to optimize the use of both soil types and enhance crop productivity. Soil type significantly impacts crop growth and yield, making it a critical factor in agricultural planning and management. By leveraging Smooth Support Vector Machine (SSVM), a powerful machine learning technique, this study aims to analyse historical crop yield data, soil properties, climate variables, and other relevant factors to identify patterns and relationships between crop production and soil types. The SSVM model will be trained and validated using datasets from both alluvial and red soil regions, enabling the classification and prediction of crop productivity in different soil types. Modern agricultural practices including soil testing, irrigation management, and the use of fertilizers and pesticides are also adopted to improve crop yields in both alluvial and red soil regions. Understanding the unique characteristics and suitability of each soil type is crucial for sustainable crop production and agricultural development in Tamil Nadu. The facts identifiedby this research will enhance comprehension of the comparative appropriateness of red and alluvial soils for various crops, providing valuable insights for agricultural decision-making and resource allocation in Tamil Nadu, India.

Keywords: Smooth Support Vector Machine; Crop Production; Alluvial Soil; Red Soil; Machine Learning; Agricultural Planning; Soil Properties; Climate Variables.

1. Introduction

Physical characteristics of soil are the ones that define its usefulness for engineering, agricultural, and environmental applications.The physical characteristics of the soil are directly associated with flow of air and heat; root penetration, water and nutrient retention as available to the plants.Chemical and biological characteristics are also influenced by physical properties.However, the unfavourable soil fertility in India has a great influence upon agriculture.The characteristics of soil vary greatly in nature [1].The parameters and characteristics can greatly alter the nature of the nation's soil structure and how it is used.Because of continuous alterations to the earth's surface and the geologic forces that from the soil, physical qualities of the soil are naturally variable.Moreover, agricultural management techniques could have induced it. Additionally, the geographical heterogeneity of soil qualities and human activity on cultivated areas have an impact on soil susceptibility to erosion [2].
The solubility of soil nutrients varies with pH, which can impact the qualities of soil, including the rate at which nutrients leach out.Agronomically speaking, interactions between soil pH and nutrients are significant since they need varied choices about the rate at which fertilizer is applied. The best growth conditions of Indian vegetable crops based on soil pH for optimum growth is by [3] and this would aid the farmers to decide on the crops to chosen for cultivation with specificity.
A soil is considered saline if it has enough soluble salts in it to negatively impact plant growth [4]. In addition to reducing crop development and having an impact on plant's water intake, salinity in the soil can lead to ion

toxicity and nutritional imbalances. Bulk density (BD) is a measure of soil health and compaction because it impacts several key processes that drive soil productivity, including infiltration, water availability, soil porosity, plant roots depth and limitations [5], nutrient availability, and microbial activity [6]. BD is influenced by the presence of organic matter, mineral density, soil texture, and how those elements are packed [7]. Additionally, the depth, restricting layers, and texture of the soil all affect the amount of water that is available. Compaction and organic matter in the soil also have an impact on its ability to hold water [8].
A vital soil nutrient, nitrogen is needed in a balanced quantity by all plants to support their growth and development. A pH (pH7) level of near neutral, nitrification, the process of microbial conversion to nitrate from NH_4^+ is faster and the nitrate intake occurs more in crops.Rhizobium, the bacterium that fixes nitrogen in legumes, coexists in symbiosis with legumes. [9] As soil acidity rises, Rhizobium's survival and activity decrease. Legumes' nodule count is influenced by the pH of the soil [10], and some species of legumes, including Mimosa and Lupins, can withstand acidic conditions. [11]
The form and availability of phosphorus (P) which is depend on soil pH, is essential for plant growth and grain production [12]. Consequently, in order to maximize the effectiveness of P fertilizers, the pH of the soil must be adjusted to an appropriate level particular to each type of soil [13]. Since crops absorb P from the soil solution, P availability depends critically on the pace at which P is replenished in other forms. The pH of the soil, the amount of phosphorus present, how the soil fixes it, and where additional phosphorus is introduced all affect how quickly the soil replenishes [14]. The availability of specific nutrients to plants and their balance among their many forms can both be impacted by pH changes in the soil [15].As there is an increase in soil pH, the availability of micronutrients like boron (B), copper (Cu), zinc (Zn), iron (Fe), manganese (Mn), and zinc (Fe) tends to decrease. Farm management methods and continuous different soil processes interactions cause variation in soil characteristics [16]. In an effort to sketch out areas of nutrient deficiency a study was conducted in Pachapalayam, Coimbatore, India regarding the spatial variability of the physical properties of the soil and macro nutrient plant growth promoters' viz., N, P and K.
The primary source of the Indian economy is agriculture.This is why India is called a land of farmers.In India, 50% of the labour force makes a living in agriculture. The first and fundamental component of agriculture is the soil [17]. However, farmers are employing the conventional approach even today.Farmers do not receive desired results from the traditional approach, which indicates that the amount of crops grown is not getting bigger. Improved soil quality is necessary to boost crop yield.Therefore soil content is being tested.Testing the soil is more significant than farming itself.Crop quality and yield are entirely dependent on the soil. Because soil testing provides information on all nutrients found in the soil, including calcium (Ca), potassium (K), and nitrogen (N), it is crucial. Drought is causing farmers in India, particularly in several areas of the state of Tamilnadu, to see a decline in crop and yield quality.They are unaware about the nutrient availability in their field. The crop using their own experience, which has a very poor success rate [18].

1.1 Characteristics of Alluvial and Red soil

Alluvial soil and red soil are two distinct types of soil found in various regions, each with its unique characteristics. Here's a comparison of the characteristics of alluvial soil and red soil:

1. **Alluvial Soil**

- Formation: Over a period of time, silt, sand, and clay carried by rivers and streams deposit to create alluvial soil. It is typically found in river valleys, floodplains, and deltas.
- Texture: Alluvial soil has a fine texture with a well-balanced mixture of sand, silt, and clay particles. It is generally loose and friable, allowing for easy cultivation.
- Fertility: Alluvial soil is highly fertile due to its mineral-rich composition derived from the parent rocks and organic matter deposited by river sediments. It can be used for a variety of crops as it has important nutrients including potassium, phosphorus, and nitrogen.
- Drainage: Alluvial soil has good drainage characteristics, allowing excess water to percolate through the soil profile easily. However, in some areas, poor drainage can lead to waterlogging during the rainy season.
- Color: The color of alluvial soil varies depending on its composition and the minerals present. It can range from light brown to dark brown or greyish-black.

2. **Red Soil**

- Formation: Red soil, also known as lateritic soil, is formed by the weathering of crystalline rocks such as granite and gneiss over millions of years. It is commonly found in upland areas and hilly regions.
- Texture: Red soil has a coarse-textured structure with a high proportion of sand particles and low clay content. It is often gritty and loose, making it prone to erosion.

- Fertility: Red soil is generally less fertile compared to alluvial soil due to its weathered and leached nature. It is deficient in nutrients such as nitrogen, phosphorus, and organic matter, requiring regular soil amendments and fertilization for agriculture.
- Drainage: Red soil has moderate to poor drainage characteristics, leading to water stagnation and runoff during heavy rainfall. It tends to become hard and compact when dry and can form crusts on the surface.
- Color: Red soil derives its characteristic reddish color from the presence of iron oxides, particularly hematite and goethite. The color can vary from deep red to brownish-red or yellowish-red.

Overall, while alluvial soil is known for its high fertility, good drainage, and suitability for a wide range of crops, red soil is characterized by its coarse texture, moderate fertility, and challenges related to drainage and nutrient deficiencies. Understanding the characteristics of each soil type is crucial for agricultural planning, soil management.The graph of productivity of agriculture in Tamilnadu has been showing a declining trend off late.Three significant clusters—designated as High, Low, and Moderate agriculture productivity index in Tamilnadu—were identified by the recent research [19]. Economic growth is reliant upon an area's net agricultural productivity, yield and production. The ratio of the index of local agricultural output to the index of all inputs utilized in farm production is referred to as productivity [20]. An area's agricultural productivity is expressed by the Agricultural Productivity Index (API). The agricultural zones are categorized into various categories based on this API.The significance of this zone classification by API stems from its ability to support policy formation. This study article uses the SVMs classification methodology with the use of data mining software version 3.3.7, one of the various ways employed for classifying all of Tamilnadu's districts based on API. This application, which was created using the Python programming language, is mostly utilized for data visualization, classification, regression, evaluation, and unsupervised classification. Orange data mining software features a strong scripting language and is incredibly user-friendly.The key objectives of this study report are (i) utilizing the Enedy'l approach to determine the Agriculture Productivity Index (API), (ii) using the aid of API index, categorize the productivity performance of different districts across the research period using SSVMs Classification data mining techniques.

2. Related Works

This section summarizes some of the important crop prediction research that has been done in the discipline of agriculture. In [21], the authors focused on the usage of data mining techniques in the sector of agriculture.Since data mining is a relatively new and developing discipline, authors have also looked into and studied the issue of agricultural productivity forecasting. Finding the ideal data models that provide high accuracy and high generality in terms of yield forecasting abilities was the primary goal of this work, according to the authors.For this, authors evaluated several data sets using various data mining techniques.

Because crop yield approaches combine a wide range of connected elements impacted by external features and non-arbitration, they are time-dependent and nonlinear by nature [22]. Traditionally, farmers used trustworthy historical data and their own past experiences to forecast crop yields and make critical production decisions. Furthermore, the development of accurate and comprehensible weed maps is made possible by remote sensing, namely satellite and aircraft multispectral scanning, photography and video, which together allow precision weed management [23]. Additionally, newly emerging machine learning (ML) algorithms are more adept at using statistical techniques to find relationships between weather and yield [24].

One method of supervised machine learning is called Support Vector Machines (SVMs).There are several instances of its application in the domain of agriculture. Provided an account of the application of SVM for precipitation reduction in climate change scenarios [25]. In order to get generalized performance and reduce the generalization error bound, SVM was employed to predict the supply and demand in the case of pulp wood. SVM was also used to provide agricultural yield forecast insights into crop response patterns related to climate conditions by offering the features contribution analysis. Support vector machines based on discretization were employed to classify agricultural datasets [26].

In [27], the Crop Selection Method (CSM) was the method that the authors offered.The authors outline the proposed approach to assist in identifying crop selection methods, increasing crop net yield rates throughout the course of seasons and assisting in achieving the highest possible level of economic growth. The authors addressed about the various contributing factors that various forecasting models which can be applied to crops. Furthermore, the authors describe machine learning and its various approaches.Crops were categorized as seasonal, whole-year, short-term, and long-term plantation crops under the proposed Crop Selection Method.

In [28], the authors proposed an innovative approach for predicting crop productivity in addition to recommending the best climate conditions to increase crop yield. In order to determine the agricultural yield per acre, multivariate polynomial regression, support vector machine regression, and random forest models were employed. The yield and weather data that the author obtained from the US Department of Agriculture are also utilized in the proposed approach. The author furthermore compared multivariate polynomial regression, support vector machine regression, and random forest with the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), median absolute error, and R-Square values.

In [29], byanalysing trends from past information, the authors employed the sliding window non-linear regression method to make recommendations for crop yield and price.Tamilnadu state's multiple districts were examined by the writers. The authors presented a technique that is intended to recommend the best crops for a farmer to plant. The system that was designed had made demand levels classification. By categorizing the dataset of changes in crop's market prices, the demand for the crops is forecasted using the demand level classification. The authors also performed the text-to-speech conversion in the proposed system.

In [30], assist Indian farmers in selecting the best crop to plant based on the season, region, and soil properties, authors have come up with an ideal system known as Agro Consultant.The authors employed machine learning algorithms such as random forests, K-NN, decision trees, and neural networks to create this kind of system. The rainfall forecaster and map visualization functionality are also included in the proposed system.

In [31], a novel framework known as the eXtensible Crop Yield Prediction Framework (XCYPF) was developed by the authors. For the purpose of crop yield forecasts, this framework offers datasets, dependent and independent factors, and crop selection options. The authors talked about how this framework might be expanded and altered. The authors forecasted crop yields for rice and sugarcane using surface temperature and rainfall data.The authors employed an innovative approach that integrates the utilization of vegetation indices.

In [32], a crop yield prognosis model (CRY) based on an adaptive cluster technique was created by the authors. Based on crop growth pattern and yield, it studies and categorizes the crop using the Bee Hive modelling approach. Also, the authors discuss the employment of Bee Hive Cluster, since it provides information on agricultural datasets that are useful in determining crop development and the Bee Hive Cluster is a heterogeneous data cluster that is kept up to date as repositories. The Bee Hive, which performed better than others, was the subject of the authors' study.The graphs were plotted to show yield variation, and the Bee Hive algorithm is used to find trends.

In [33], In order to enable Indian farmers and the agricultural sector to obtain important information about agriculture, authors developed a cloud-based agricultural framework. This approach offers agricultural yield predictions and soil classification.

3. Research Gap

Based on the existing literature and the proposed study on the SVM-based comparative analysis of crop production in alluvial soil and red soil regions of Tamil Nadu, India, there are several potential research gaps that could be addressed:

1. **Limited Application of Machine Learning Techniques:** While traditional statistical methods have been used to analyse crop production and soil types in Tamil Nadu, the use of cutting-edge machine learning methods, such as Support Vector Machines (SVMs), is lacking. Exploring the potential of SVMs in analysing agricultural data could provide new insights and complement existing research.
2. **Integration of Multiple Data Sources:** Many studies focus solely on soil characteristics or crop productivity without considering the interplay between various factors such as climate, agricultural practices, and socio-economic variables. There is a gap in integrating multiple data sources to comprehensively analyze crop production dynamics in different soil regions.
3. **Scalability and Generalizability:** Some research on crop production in Tamil Nadu may lack scalability and generalizability due to limited sample sizes or localized studies. Addressing this gap by using robust datasets and rigorous methodologies can improve the reliability and applicability of the findings.
4. **Temporal Analysis:** Many studies provide a snapshot of crop production and soil characteristics at a specific point in time, but there is a gap in conducting temporal analysis to assess changes and trends over time. Longitudinal studies can help identify temporal patterns and drivers of change in crop production dynamics.
5. **Policy Implications and Stakeholder Engagement**: While research on crop production and soil types provides valuable insights, there is often a gap in translating these findings into actionable recommendations for policymakers, agricultural stakeholders, and farmers. Bridging this gap requires engaging with stakeholders throughout the research process and highlighting the policy implications of the findings.

NATURALISTA CAMPANO
ISSN: 1827-7160
Volume 28 Issue 1, 2024

6. Addressing these research gaps can enhance our understanding of crop production dynamics in alluvial and red soil regions of Tamil Nadu, India, and contribute to more effective agricultural planning, resource management, and policy formulation.

4. Proposed Method

The process flow of proposed system as shown in Figure 1. It's important to note that this process requires expertise in machine learning, data analysis, and agricultural sciences to ensure accurate interpretation and meaningful insights from the SVM model results. Furthermore, the dependability of the analysis is greatly influenced by the representativeness and quality of the data.
Steps to follow using proposed method

1. Data Collection: Gather data on crop production in both alluvial soil and red soil regions of Tamil Nadu. This data should include variables such as crop type, yield, soil type, weather conditions, agricultural practices, etc.
2. Data Pre-processing: Cleaning the data to get rid of any flaws or inconsistencies. This could mean addressing outliers, taking care of missing values, and data normalization.
3. Feature Selection/Extraction: Identify relevant features that may influence crop production in each soil type. This could include soil characteristics, climate data, agricultural inputs, etc.
4. Training and Testing Data Split: Separate the training and testing sets from the dataset. The SVM model is trained on the training set, and its performance is assessed on the testing set.
5. Comparative Analysis: Compare the performance of the SVM model for predicting crop production in alluvial soil versus red soil regions. Analyse factors such as accuracy, sensitivity to different crops, and any insights gained from the model's decision boundaries.
6. Interpretation: Interpret the results of the comparative analysis. Identify any significant differences in crop production between the two soil types and discuss possible reasons for these differences based on the SSVM model's findings.
7. Validation: Validate the results of the SSVM analysis using additional techniques or by comparing them with existing literature on crop production in similar regions.

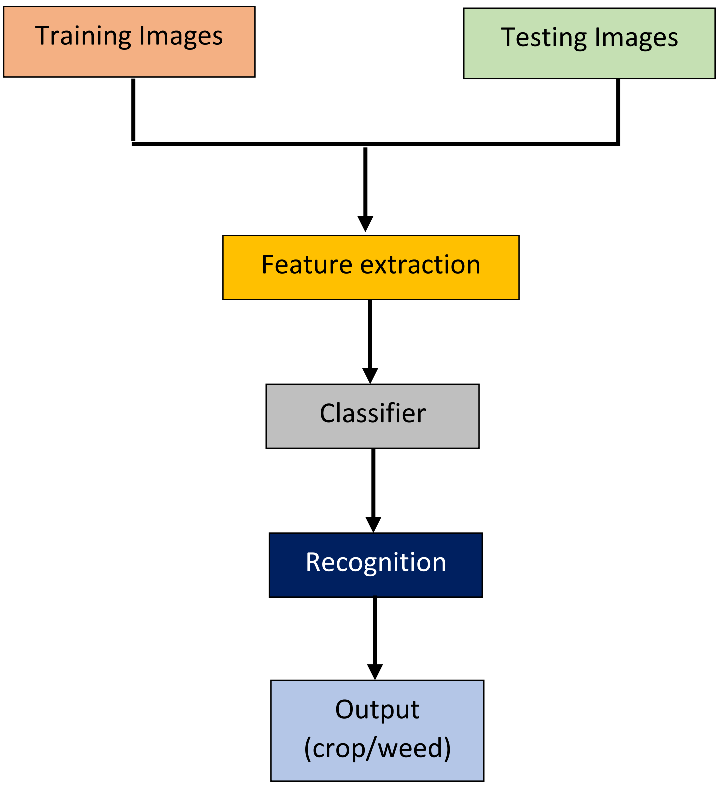


Figure 1: Proposed system processing steps

3.1 Data Collection

Acquiring relevant datasets is crucial for conducting the SSVM-based comparative analysis of crop production in alluvial soil and red soil regions of Tamil Nadu, India. Here are some potential sources and types of datasets may consider

- Crop Production Data: Obtain data on crop yields, acreage, production quantities, and harvest timings for various crops cultivated in alluvial and red soil regions of Tamil Nadu. This data may be available from government agricultural departments, agricultural research institutes, or agricultural surveys.
- Soil Characteristics Data: Gather information on soil properties such as pH levels, organic matter content, nutrient composition, texture, and water retention capacity for alluvial and red soil regions. Soil data can be obtained from soil surveys, soil testing laboratories, or research institutions.
- Climate and Weather Data: Collect historical climate and weather information on the sun's radiation, temperature, humidity, rainfall, and evapotranspiration for the study area. This data can be sourced from meteorological departments, weather stations, or online repositories.
- Agricultural Practices Data: Access data on agricultural practices and inputs such as irrigation methods, fertilizer usage, pesticide application, crop rotation patterns, and cropping intensity in alluvial and red soil regions. Agricultural extension services, farm surveys, or research studies may provide this information.
- Satellite Imagery: Utilize remote sensing data from satellites to assess vegetation indices, land cover types, and crop health indicators in alluvial and red soil regions. Satellite imagery can be obtained from platforms such as NASA Earth Observing System Data and Information System (EOSDIS) or commercial satellite providers.
- Geospatial Data: Incorporate geospatial datasets such as land use/land cover maps, elevation models, and hydrological layers to analyse the spatial distribution of crops and soil types in Tamil Nadu. Geospatial data may be available from government agencies, research institutions, or online repositories.
- Socio-Economic Data: Consider socio-economic indicators such as farm size, household income, agricultural labour availability, market access, and infrastructure development in alluvial and red soil regions. Socio-economic data can be obtained from census records, household surveys, or socioeconomic studies.
- Open Data Portals: Explore open data portals such as the Indian Ministry of Agriculture and Farmers Welfare's data portal, Indian Council of Agricultural Research (ICAR) data repositories, or platforms like Data.gov.in for publicly available datasets related to agriculture in Tamil Nadu.
- Research Studies: Review existing research studies, academic publications, and reports on crop production, soil types, and agricultural practices in Tamil Nadu. Some researchers may share their datasets as supplementary materials or upon request.

Ensure to assess the quality, completeness, and reliability of the datasets before using them for analysis. Combining multiple datasets from diverse sources can provide a comprehensive understanding of crop production dynamics in alluvial and red soil regions, facilitating informed decision-making and policy formulation.
https://www.kaggle.com/datasets/aishu200023/tamilnadu-cropproduction.
The information is gathered from the website of the Soil Health Card. Twenty-three crops are included in the data viz., rice, corn, beans, wheat, sugarcane, tea, mulberries, etc. As detailed in Table I, it has 22 properties, including Ph, Mn, Fe, EC, OC, B, Zn, N, P, K, and so on. The information is collected for several Tamil Nadu districts, viz., Tirunelveli, Madurai, Kanyakumari, and Virudhunagar.
Example: The software WEKA (Waikato Environment for Knowledge Analysis) is used to implement the proposed predictive model. Farmers growing paddy along the Thamirabarani river basin provided the agricultural data set, which includes characteristics such crop variety, land type, soil fertility, and seed quality. This authentic dataset was created by utilizing 200 question given in surveys to farmers who were cultivating rice fields in the districts of Tirunelveli and Tuticorin, which are watered by the Thamirabarani River.

Regions in Tamilnadu

In Tamil Nadu, India, alluvial soil and red soil regions are distributed across various districts shown in Figure 2.

Alluvial Soil Regions

- Thanjavur District: Located in the Cauvery delta region, Thanjavur district is known for its fertile alluvial soil, particularly suitable for paddy cultivation. The district is renowned for its rice production and is often referred to as the "Rice Bowl of Tamil Nadu."

NATURALISTA CAMPANO
ISSN: 1827-7160
Volume 28 Issue 1, 2024

- Tiruchirappalli District: Situated along the banks of the Cauvery River, Tiruchirappalli (Trichy) district features fertile alluvial plains conducive to agriculture. Paddy cultivation is prevalent in this region, along with the cultivation of other crops such as sugarcane and pulses.
- Nagapattinam District: Another district located in the Cauvery delta region, Nagapattinam district has extensive areas of alluvial soil suitable for rice cultivation. The district is known for its coastal agriculture and is a significant contributor to Tamil Nadu's rice production.

Red Soil Regions

- Coimbatore District: Located in western Tamil Nadu, Coimbatore district features extensive areas of red soil. The district is well-known for its wide range of agricultural pursuits, which include the raising of commodities such as millets, pulses, oilseeds, cotton, and spices in the red soil regions.

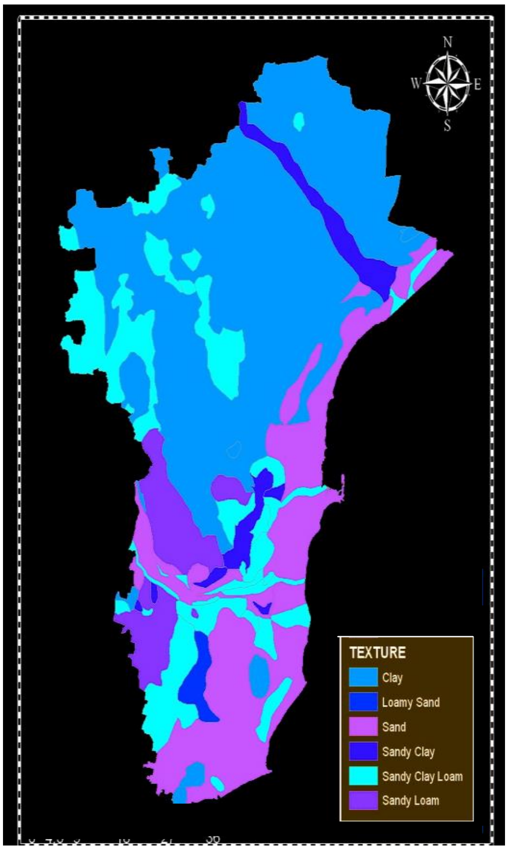


Figure 2: Alluvial and Red soil regions in Tamilnadu

- Salem District: Salem district is renowned for its red soil areas, particularly in the Shevaroy and Yercaud hills. Agriculture in Salem district includes the cultivation of crops like millets, pulses, oilseeds, and fruits such as guava and pomegranate in the red soil regions.
- Dindigul District: Situated in central Tamil Nadu, Dindigul district has significant areas of red soil, particularly in the foothills of the Western Ghats. Agriculture in Dindigul district includes the cultivation of crops such as millets, pulses, oilseeds, and spices in the red soil regions.

These are some of the prominent districts in Tamil Nadu known for their alluvial soil and red soil characteristics, each contributing to the state's diverse agricultural landscape and crop production.

3.2 Data pre-processing

One method to raise the calibre of data that is sent to the mining process is pre-processing. High-quality input data will yield fascinating and very helpful insights. There are four main methods for doing data pre-processing in the proposed system: (i) data cleaning, (ii) attribute selection, (iii) transformation, and (iv) integration.

3.2.1 Data cleaning

The process of replacing noisy, inconsistent, and incomplete data is known as data cleansing. There are missing or null values for several attributes. Null values are eliminated during cleaning in order to produce high-quality knowledge.

The average value of attributes with a small number of null values—like the amount of fertilizer—is updated. This is accomplished by utilising a filter in WEKA - filters.unsupervised.attributes.replacemissingvalue. Attributes with too many null values (e.g., type of pesticide) are removed using a filter in WEKA - filters. unsupervised.attributesremove.

3.2.2. Attribute selection

This procedure identifies the attributes that are more important to the mining process. The process of selecting attributes involves removing redundant and unnecessary attributes. In WEKA, attributes are sorted according to information acquired using an attribute selection filter. By virtue of attribute selection, characteristics with extremely low information gain scores—like seed quality, irrigation technique, and crop spacing—are eliminated.Table 1 contains a list of the input parameters that were used in this study.

3.2.3. Transformation

Transformation is the process of transforming the data into a format appropriate for a mining task. In WEKA using the filter unsupervised.attributes.numerictonominal option, numerical value attributes (e.g, temperature) are transferred into categorical attributes.

3.2.4. Data integration

The agricultural data gathered from the questionnaires which were filled up by the farmers in several places throughout the Thamirabharani river basin are complied.

Table 1: Select attributes list

S.No	Attribute type	Values
1	Soil types	(Black, alluvial, red, clay)
2	Crop types	(rice, sugarcane, cotton,various pulses, vegetables, millets, fruits)
3	Quality of seeds	(Good, Medium, Low, Poor)
4	Rate of seed	(<=25, 26-30, >30)
5	Season type	(Autumn, Winter, Summer, Rainy)
6	Fertilizer per acre	(1 kg per Acre)
7	Rainfall	(Good, medium, low, poor)
8	Method for preparing land	(Tillage, Planting)
9	Sowing process	(mechanical, manual)
10	Type of fertilizer	(Complex, potassium, urea, dung, sulphide)
11	Pesticides amount used per acre	(In Liters per kg)
12	Rotation of crops	(All crops used in alluvial and red soils)
13	Manure added naturally	(No, Yes)
14	Temperature type	(High, medium, low)
15	Yield per acre	(Good, Bad
16	Fertility of soil	(Nitrogen, potassium-high, medium, low,

		phosphorus

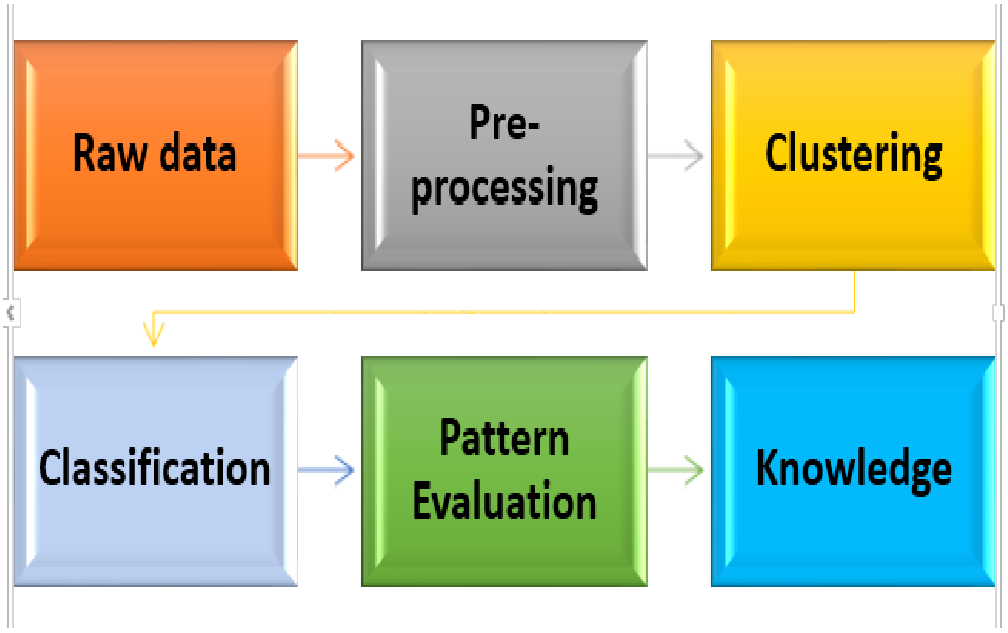


Figure 3: Framework for proposed system

3.2.5 Parameters used

Every dataset that was used in this study came from the Indian government's publicly available records. Only a small number of significant variables with the greatest influence on agricultural productivity were chosen for the current study from the enormous initial dataset. The parameters chosen for the research are displayed in Figure 3 below.

- Precipitation (mm): The monthly mean precipitation of that year for a certain district was used to compute the total precipitation for the Kharif season (June to November) for each year of every district.
- Minimum, Average, Maximum Temperature (degree Celsius): Variations in temperature will ultimately affect crop productivity. For the purposes of this research, the lowest, average, and maximum temperatures for each district's every year were taken into account.
- Reference Crop Evapotranspiration (mm): Every district's reference crop evapotranspiration was determined using the monthly mean for that particular year's Kharif season.
- Area (Hectares): For this study, the area under rice cultivation in each district of the state of Tamilnadu from June to November of each year during the Kharif season was taken into account.
- Production (Tonnes): The current study took into account the rice output for the aforementioned planted area during the Kharif season (June to November) in each of the districts of Tamilnadu state that were chosen.
- Yield (Tonnes/Hectare): For the purpose of this research, the calculated yield was taken into account for each year of the districts selected from the State of Tamilnadu which was based upon the rice production and area under rice cultivation during Kharif season.

3.3 Feature Extraction/ Selection

In the majority of image processing methods that entail weed detection, feature extraction is regarded as an essential phase. The ability to distinguish between items of different coloration, such as soil and crop, can be achieved by using the color feature in photographs.On the opposite, there are two categories for shape features: region descriptors and boundary descriptors.Texture is regarded as the primary optical feature that embodies the intrinsic external characteristics of an object and how they relate to the local environment. With just their textures and no further information, sparse objects in an image can be identified with absolute certainty. Texture gives information on the orderly arrangement of an image's external look. There are two ways for obtaining surface features: Structured Method and Statistical Method. Artificial textures can be accurately rendered using this technology. In contrast, a statistical approach takes into account an image's texture from a different perspective when determining the potency distribution in a given area.

3.4 Analysis of SSVM by statistical methods

The following steps discuss about the data mining SSVMs algorithms. A classification method called SSVM maximizes the margin between examples of distinct classes by using a hyper plane to divide the attribute space shown in Figure 4. This method produces extremely high prediction performance outcomes often. A graphical user interface is what this widget is by itself. The SSVMs Classification will be carried out by the following algorithm:

Algorithm SSVM

Step 1: After choosing the file widget, database in .tax and .xls format could be loaded.

Step 2: In order to connect to a SVM widget, first a data table widget needs to be connected to file widget, which will then be connected to SCM widget.

Step 3:

To achieve an input matrix of production and in an unit area, the area of all the crops is calculated and assigned as Y, then divided by production of the selected crops in the entire area (Yn)

Step 4: The second section is calculated to provide an input matrix of the chosen crops' cultivation areas within the districts (T), which is then divided by the zone's total selected crop cultivation areas.

Step 5: The API is calculated using step 1 divided by step 2 and multiplied by 100. The formula for API is :

$API = \frac{Y}{Y_n} \div \frac{T}{T_n} \times 100$ (1)

Step 6. It is possible to give the learner a name that will show under it in other widgets.SVM‖- will be the default name.

Step 7.Setting the classification and test error settings comes as the next step. The foundation of C-SVM and v-SVM differs in how they minimize the error function.

Step 8. The function that fits the maximum-margin hyper plane by transforming attribute space into a new feature space is called a kernel, and it is this function that allows the algorithm to build non-linear classifiers with Polynomial, RBF, and Sigmoid kernels. This is covered in the next block of options.In addition to their names, the functions that define the kernel are presented. The constant values for the kernel function, the degree of the kernel, and the gamma constant in the kernel function are all set to 0 by default.

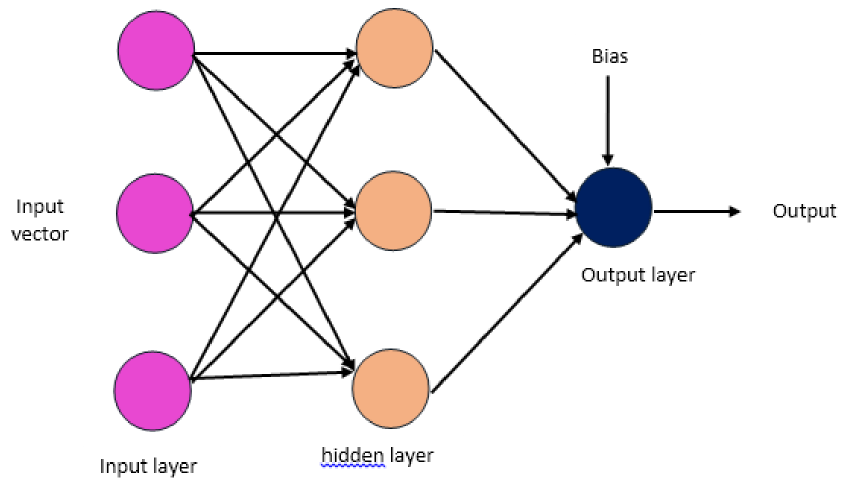


Figure 4: SSVM Architecture

Step 9: To complete the process, double click each of the following widgets in the same order:file widget- data table widget- Mosaic graph widget- SVM widget. When these widgets are executed, their classification tables and mosaic graphs are displayed for the duration of the study is shown n Figure 5.

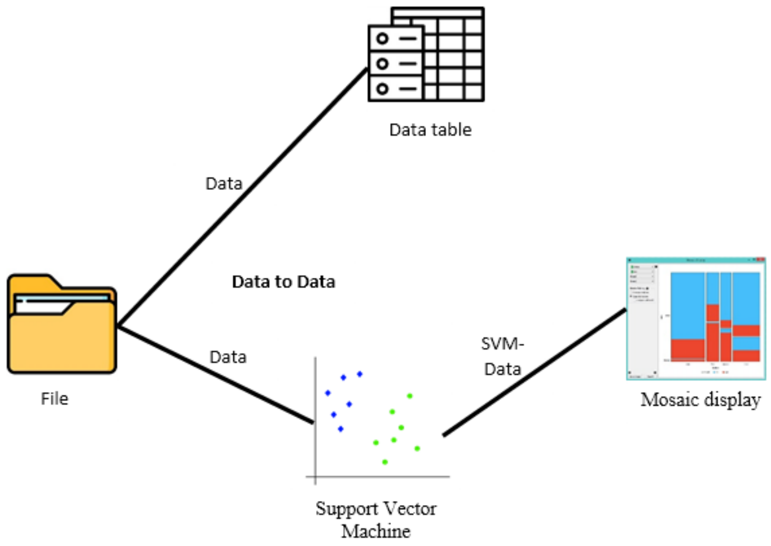


Figure 5: Flow diagram of SSVM

3.5 Training and Testing the model

Divide the dataset into sets for testing and training. The SSVM model will be trained in the training set, and its efficacy will be assessed on the testing set. Choose an appropriate SSVM implementation or library for your analysis. There are several libraries available in Python, such as scikit-learn, LIBSVM, and SVM light, which support SSVM.Set up the SSVM model with the proper initialization parameters, such as the kernel coefficient (gamma), regularisation parameter (C), and kernel function (polynomial, radial basis function, linear). Utilising the training dataset, SSVM model may be trained. This model will learn to predict crop production based on the selected features and target variable (crop yields).Configure the selected kernel and hyper parameters to lessen the error between the projected and actual crop production values, then fit the SSVM model to the training set. Utilising the evaluation dataset, assess the efficiency of the trained SSVM model. For regression tasks, the Mean Absolute Error (MAE), R-squared (R^2) score and Mean Squared Error (MSE) are often used evaluation metrics. To evaluate the accuracy and dependability of the model, compare its forecasts with the actual values of crop output.Validate the SSVM model's performance using techniques like k-fold cross-validation to ensure its robustness and generalization to unseen data. Interpret the SSVM model's results to gain insights into the factors influencing crop production in Tamil Nadu. Analyze the model's decision boundaries, support vectors, and feature importance to understand how different features contribute to the prediction.

3.6 Classification of Soil

The basic procedure employed in this proposed study for the prediction and classification of soil characteristics is shown in Figure 6. Samples are gathered using IoT-based smart farming for cross-validation and training, and information about the samples is sent instantly between farms and with farmers**.**75% of the collected specimens are utilized and the other 25% are utilized in examination after they are randomly reassembled**.** In cross-validation, ten steps are followed, with 10% of the information used for evaluation and the rest used for the data for training in each step.The data is used to train the extreme learning method (ELM), a quick learning categorization technique, to detect the micronutrients in the soil. The training processes and the overall amount of nodes that are concealed are the parameters that are engaged.The best parameters used to test data are initially chosen from the training set.The final result of the test is arrived and the micronutrients are identified after an average of 10 attempts.By allocating values of 50 and 150, respectively, to the hidden neuron parameter, the ELM classifier in this study is able to classify pH as well as detect soil nutrients. Consequently, tuning is done within [10,150] and [10,200], correspondingly. In order to optimize the model, similar activation parameters are applied, including the triangular basis, hyperbolic tangent, hard limit, Gaussian radial and sine-squared basis.

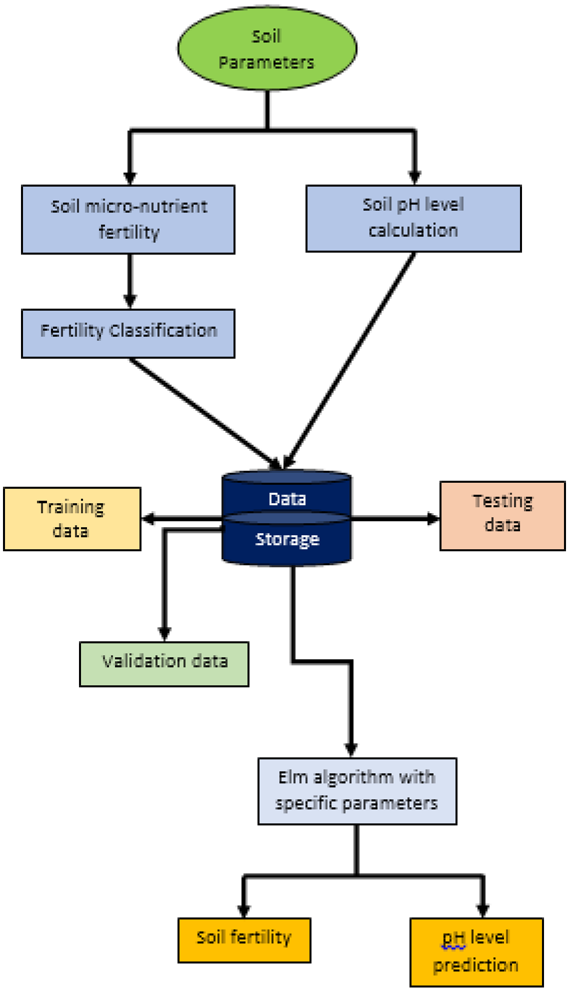


Figure 6: Proposed model to predict pH level and fertility of soil

4. Results and Discussions

Four districts in Tamil Nadu have their soil samples and the available micronutrients grouped based on the analysis that was done.

Salem. The soil's general texture has an alkaline reaction of more than 8.2 and a neutral salinity. The micronutrients obtained by DTPA were found to be adequate in respect of Cu and Fe, however, the soil exhibited deficiencies for Mn, B, and Zn.

Coimbatore. The major portion of the Coimbatore district's soils are shown to have a satisfactory sulphur level with >12 ppm.

Tirupur. The majority of the soils in the Tirupur district are found to have low sulphur contents.The soil's general texture has an alkaline reaction of more than 8.2 and a neutral salinity. The micronutrients acquired by DTPA were found to be adequate in respect of Cu, B, and Mn,; however, the soil exhibited deficiencies in respect of Fe and Zn. A neutral salinity of less than 1 mmhos/cm and a relatively high pH of 8.24 were also exhibited by the soil.

Erode. The bulk of the soils in the Erode district have been discovered to have a high sulphur concentration. The micronutrients acquired by DTPA were found to be adequate in Cu and Fe, and the soil exhibited deficiencies for Mn, B, and Zn.

Tamil Nadu's soils are red in colour and have a normal electrical conductivity, according to the study that was conducted. Of the samples, 75% had medium sulphur levels and the remaining 25% had low ones. Conversely, it is discovered that DTPA extractable micronutrients such magnesium and boron are sufficient in 20% and 60% of the samples and low in 80% and 40% of the samples, respectively.Nonetheless, the availability of Cu and Zn was determined to be either sufficient or inadequate in all 100 samples. Table 2 presents the total soil

composition as well as the corresponding metrics for the four districts that were the subject of the study: Erode, Tirupur, Coimbatore, and Salem.
The cross-validated accuracy score for the hidden neurons that are utilised to determine high accuracy is shown in Figure 7.It has been determined that the ideal pH classification for soil nutrient classification is 150.

Table 2: Soil composition of Micronutrients

Parameters	Salem	Coimbatore	Tirupur	Erode
Cu (ppm)	1.33	1.37	1.13	1.43
Mn (ppm)	2.10	6.83	2.3	8.8
Fe (ppm)	2.35	2.97	2.65	3.47
B (ppm)	0.36	0.20	0.48	0.36
Zn (ppm)	0.46	0.37	0	0.8
S (ppm)	10.13	18.66	38.07	47.74
K (kg/ha)	276.84	295	269	399
P (kg/ha)	33.7	39	15	17.92
N (kg/ha)	523	383	395	799
OC (%)	0.6	0.43	0.39	1.7
EC (millimhos/cm)	0.19	0.10	0.083	0.3
pH	8.4	8.4	8.12	8.5

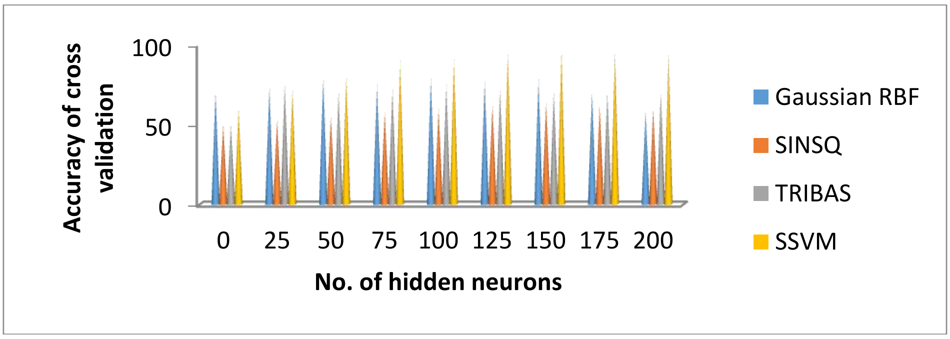
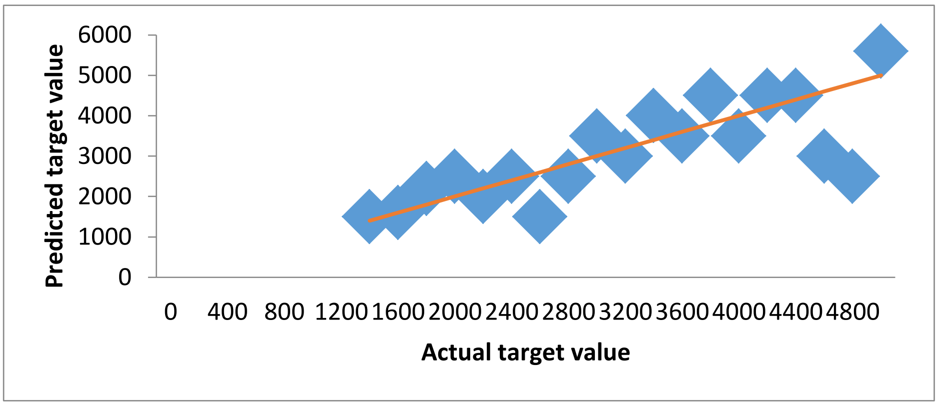
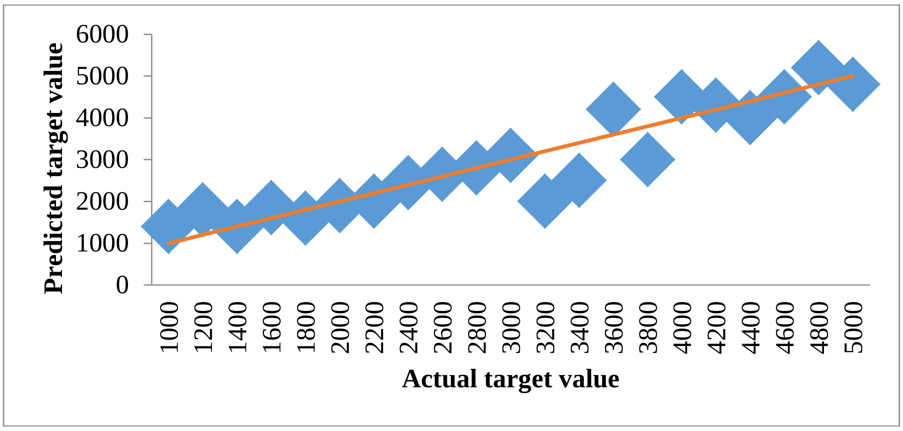


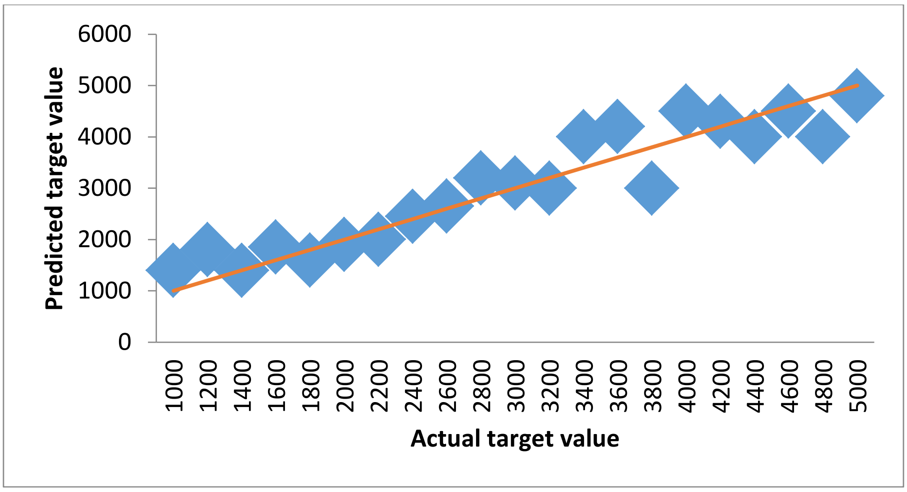
Figure 7: Comparison of hidden layers and cross validation to identify the soil pH content



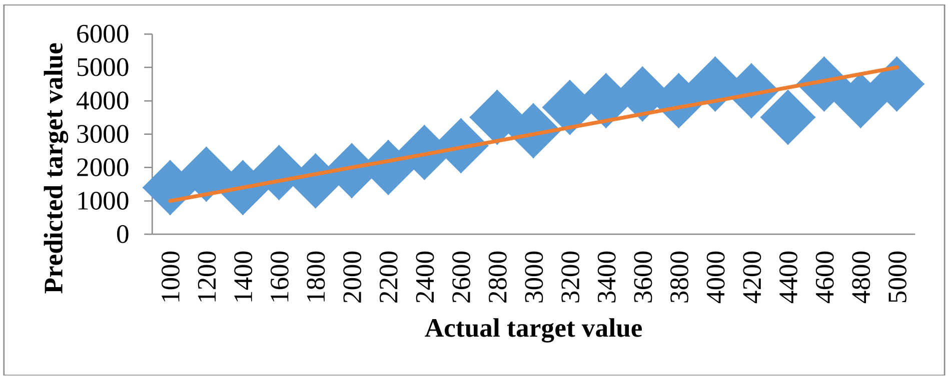
(a)


Predicted target value
6000
5000
4000
3000
2000
1000
0
1000 1200 1400 1600 1800 2000 2200 2400 2600 2800 3000 3200 3400 3600 3800 4000 4200 4400 4600 4800 5000
Actual target value

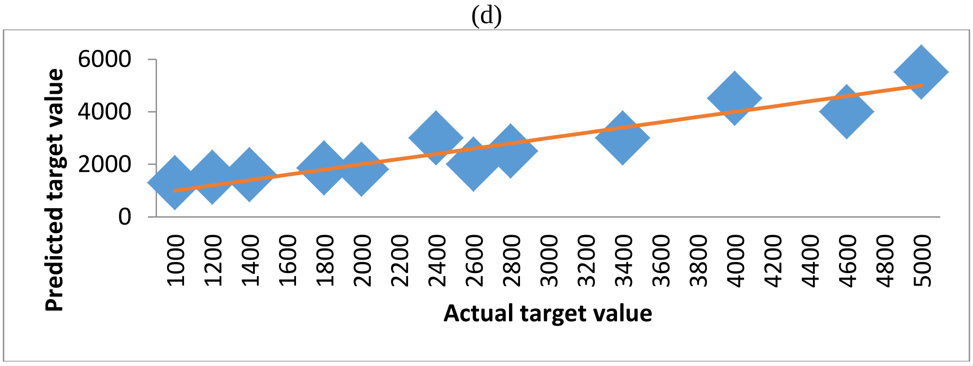
(b)


Predicted target value
6000
5000
4000
3000
2000
1000
0
1000 1200 1400 1600 1800 2000 2200 2400 2600 2800 3000 3200 3400 3600 3800 4000 4200 4400 4600 4800 5000
Actual target value

(c)


Predicted target value
6000
5000
4000
3000
2000
1000
0
1000 1200 1400 1600 1800 2000 2200 2400 2600 2800 3000 3200 3400 3600 3800 4000 4200 4400 4600 4800 5000
Actual target value

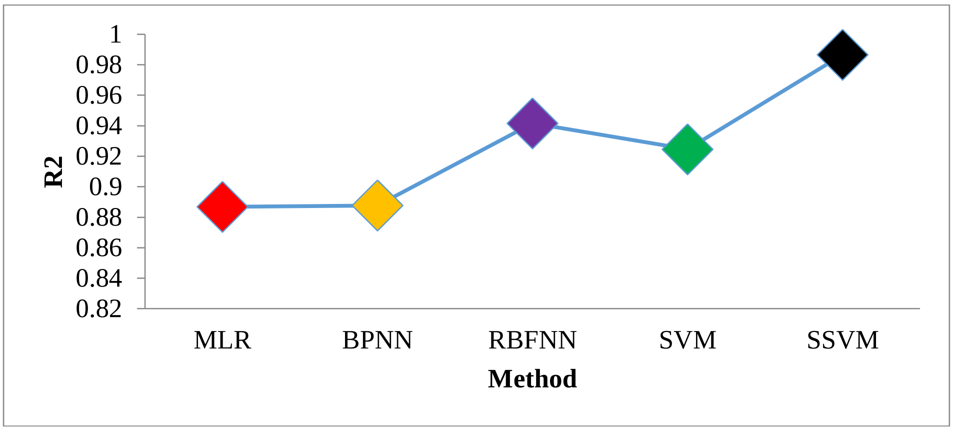
(d)


Predicted target value
6000
4000
2000
0
1000 1200 1400 1600 1800 2000 2200 2400 2600 2800 3000 3200 3400 3600 3800 4000 4200 4400 4600 4800 5000
Actual target value

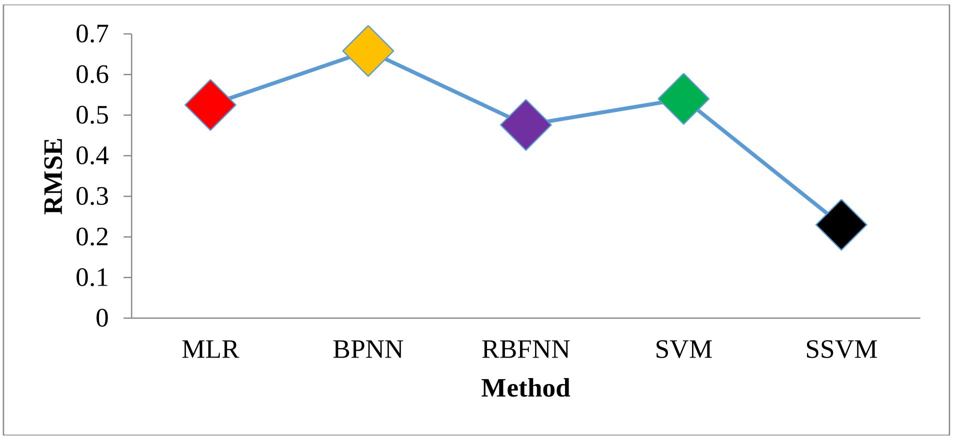
(e)

Figure 8: Crop yield between actual and predicted value

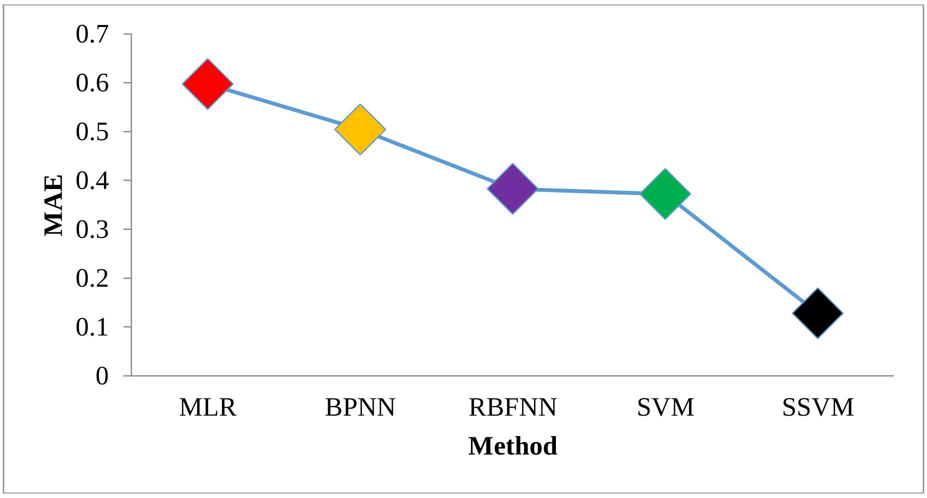
Figure 8 shows the difference in crop yield (Q/ha) between the measured and predicted data. Machine learning techniques have been shown in several studies to be capable of predicting paddy output. For a consistent crop yield, the prediction accuracy must be improved consistently. Also, every model produces a graph (Figure 8) that compares the predicted crop yield with the actual yield data.Based on the examples, all the metrics are evaluated with the aforementioned formulae. The accuracy of the algorithms under consideration is then tested by evaluating the metrics between the original and projected crop yield.



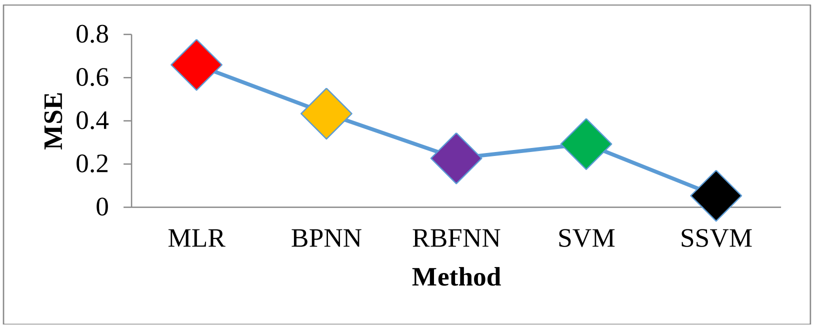
(a)



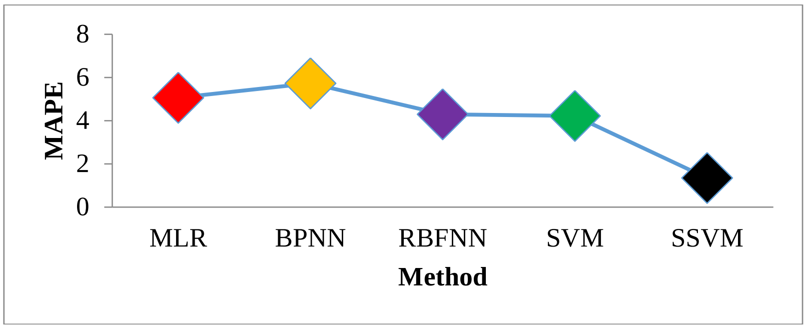
(b)



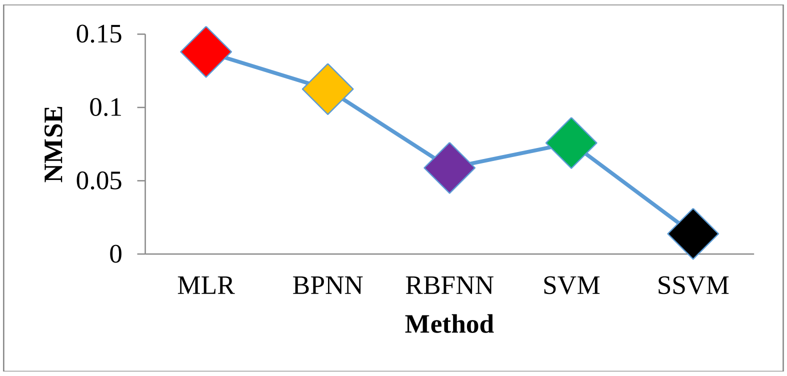
(c)



(d)



(e)



(f)

Figure 9: Comparison of performance measuresof different methods

The separate individual metrics R^2, RMSE, MAE, MSE, MAPE, CV, and NMSE are shown in Figure 9 in order to make the comparison study of different algorithms in a simple manner. A closer to unity, i.e, greater R^2represents high accuracy and lowerlower RMSE, MAE, MSE, MAPE, CV, and NMSE as stated in pre-paras.This claim is supported by the fact that the SSVM algorithm outperformed existingANN and statistical techniques.

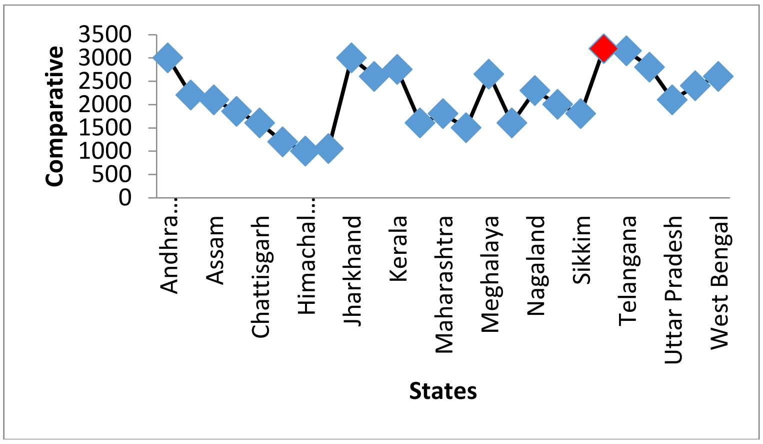


Figure 10: Comparative study of yield among India states

Additionally, the chosen location's absolute yield has been compared with that of other Indian state and Figure 10 depicts the entire comparison example. The state of Tamilnadu was found to have produced the maximum yield, at roughly 3191 kg/ha.This is because of the ideal conditions existing in the state, with a mean temperature of 28 °C, a higher average rainfall of 464 mm for a period of 3 months, and a pH of roughly 6.9. Furthermore, a variety of alluvial soil types that are appropriate for paddy farming can be found in Tamilnadu's paddy production areas. The machine learning model's predicted numbers are almost equal to Tamilnadu yield's absolute yield, while the accuracy varies depending on how well effective each algorithm works. As previously mentioned, the SSVM model's accuracy outperforms other chosen machine learning models in terms of scalability.

5. Conclusions

The analysis highlights the distinct characteristics of alluvial and red soil regions in Tamil Nadu. Alluvial soil regions, such as those found in the Cauvery delta, exhibit higher fertility and better water retention capacity compared to red soil regions, which are often characterized by lower fertility and coarser texture.The SSVM analysis demonstrates that soil type significantly influences crop production in Tamil Nadu. Crops cultivated in alluvial soil regions, such as rice and sugarcane, generally exhibit higher yields compared to those grown in red soil regions, where crops like millets and pulses are more prevalent.In both alluvial and red soil plantations, soil properties such as pH, organic matter content, and nutrient levels are found to be important determinants of crop yield. Additionally, climate variables such as temperature, rainfall patterns, and humidity play a significant role in determining crop yields in different regions. The SSVM models developed for the comparative analysis demonstrate reliable performance in predicting crop production based on soil and climate variables. Evaluation metrics such as accuracy, precision, and recall indicate the models' effectiveness in capturing the complex relationships between soil characteristics, climate factors, and crop yields.The findings have practical implications for agricultural practices and resource management in Tamil Nadu. Tailoring soil management strategies, irrigation practices, and crop selection to suit the specific characteristics of alluvial and red soil regions can help optimize crop productivity and enhance agricultural sustainability.The comparative analysis provides valuable insights for policymakers and agricultural stakeholders in formulating evidence-based policies and interventions to support sustainable agricultural development in Tamil Nadu. Future research directions may include exploring additional factors influencing crop production and conducting longitudinal studies to monitor agricultural dynamics over time.

Future Directions

It may include expanding the scope of the analysis to incorporate additional factors influencing crop production, such as socio-economic factors, pest and disease management, and technological innovations.
Longitudinal studies and continuous monitoring of crop production dynamics can provide valuable insights into the resilience and adaptability of agricultural systems to changing environmental conditions.

References

[1] Padmapriya, J., & Sasilatha, T. (2023). Deep learning based multi-labelled soil classification and empirical estimation toward sustainable agriculture. Engineering Applications of Artificial Intelligence, 119, 105690.

[2] Rejith, R. G., Sundararajan, M., Gnanappazham, L., Seenipandi, K., & Ramaswamy, S. (2022). GIS-based machine learning algorithms for mapping beach placer deposits in the southwest coast of India using Landsat-8 OLI images. Journal of Applied Remote Sensing, 16(1), 012011-012011.
[3] Paneerselvam, B., Ravichandran, N., Li, P., Thomas, M., Charoenlerkthawin, W., & Bidorn, B. (2023). Machine learning approach to evaluate the groundwater quality and human health risk for sustainable drinking and irrigation purposes in South India. Chemosphere, 139228.
[4] Sivasubramaniyan, T., & RajaPerumal, R. N. (2024). Original Research Article Identifying land use land cover dynamics using machine learning method and GIS approach in Karaivetti, Tamil Nadu. Journal of Autonomous Intelligence, 7(3).
[5] Sindhu, G., & Rathi, S. Inquisition of Landslide Hazards in Tirupattur District, South India Using Geospatial and Machine Learning Techniques.
[6] Vellaikannu, A., Palaniraj, U., Karthikeyan, S., Senapathi, V., Viswanathan, P. M., & Sekar, S. (2021). Identification of groundwater potential zones using geospatial approach in Sivagangai district, South India. Arabian Journal of Geosciences, 14(1), 8.
[7] Prabavathi, R., & Chelliah, B. J. (2022). A Comprehensive Review on Machine Learning Approaches for Yield Prediction Using Essential Soil Nutrients. Universal Journal of Agricultural Research, 10(3), 288-303.
[8] Wankhede, D. S. (2021). Analysis and prediction of soil nutrients pH, N, P, K for crop using machine learning classifier: a review. In International Conference on Mobile Computing and Sustainable Informatics: ICMCSI 2020 (pp. 111-121). Springer International Publishing.
[9] Shwetha, G. N., & Jairam, B. G. (2023, February). Continuous Real Time Sensing and Estimation of In-Situ Soil Macronutrients. In Proceedings of the International Conference on Paradigms of Computing, Communication and Data Sciences: PCCDS 2022 (pp. 573-589). Singapore: Springer Nature Singapore.
[10] Sumathi, P., Karthikeyan, V. V., Kavitha, M. S., & Karthik, S. (2023). Improved Soil Quality Prediction Model Using Deep Learning for Smart Agriculture Systems. Computer Systems Science & Engineering, 45(2).
[11] Haritha, J., Valarmathi, R. S., & Kalamani, M. (2022). Soil Urea Analysis Using Mid-Infrared Spectroscopy and Machine Learning. Intelligent Automation and Soft Computing, 32(3), 1867-1880.
[12] Meenakshi, M., & Naresh, R. (2023). Smart Agriculture: Soil Aggregate Stability Classification for Damaged Crops in India. International Journal on Advanced Science, Engineering & Information Technology, 13(2).
[13] Rao, R. R., Reddy, U. S., & Reddy, C. R. M. (2023, June). Soil Data Analytics using Machine Learning Techniques–A Survey. In 2023 3rd International Conference on Intelligent Technologies (CONIT) (pp. 1-7). IEEE.
[14] Lad, A. M., Bharathi, K. M., Saravanan, B. A., & Karthik, R. (2022). Factors affecting agriculture and estimation of crop yield using supervised learning algorithms. Materials Today: Proceedings, 62, 4629-4634.
[15] Dhoke, P., Purohit, N., & Lokulwar, P. P. (2023, July). Agroecology: Intelligent Implementation of Digital Farming Using machine Learning & Deep Learning. In 2023 World Conference on Communication & Computing (WCONF) (pp. 1-7). IEEE.
[16] Sreedevi, B., Mohanraj, G., Revathy, J., & Roobini, R. (2022, January). Agri Brilliance-A Farm Log Rental Service Platform with Crop and Disease Management Using Machine Learning Techniques. In 2022 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI) (pp. 1-7). IEEE.
[17] Chandrakumar, T., & Sakthipriya, D. (2023). A Study on Usage of Agricultural Engineering Equipment for Various Crops and Yields in South Tamilnadu. International Journal of Social Ecology and Sustainable Development (IJSESD), 14(1), 1-15.
[18] Puneeth, B. R., & Nethravathi, P. S. (2021). A literature review of the detection and categorization of various arecanut diseases using image processing and machine learning approaches. International Journal of Applied Engineering and Management Letters (IJAEML), 5(2), 183-204.
[19] Dey, B., Abir, K. A. M., Ahmed, R., Salam, M. A., Redowan, M., Miah, M. D., & Iqbal, M. A. (2023). Monitoring groundwater potential dynamics of north-eastern Bengal Basin in Bangladesh using AHP-Machine learning approaches. Ecological Indicators, 154, 110886.
[20] Vijayalakshmi, R., Thangamani, M., Ganthimathi, M., Ranjitha, M., & Malarkodi, P. (2021, August). An automatic procedure for crop mapping using agricultural monitoring. In Journal of Physics: Conference Series (Vol. 1950, No. 1, p. 012053). IOP Publishing.

[21] Shaikh, S. G., Kumar, B. S., Narang, G., & Pachpor, N. N. (2024). Original Research Article Hybrid machine learning method for classification and recommendation of vector-borne disease. Journal of Autonomous Intelligence, 7(2).
[22] Liang, R., Dai, K., Xu, Q., Pirasteh, S., Li, Z., Li, T., ... & Fan, X. (2024). Observation and Geoinformation. International Journal of Applied Earth Observation and Geoinformation, 127, 103654.
[23] Sasikala, S., & Bharatula, S. D. (2024). IoT-Based Smart Irrigation System Based Adaptive Radial Deep Neural Network (ARDNN) Algorithm Applicable for Various Agricultural Production. International Journal of Intelligent Systems and Applications in Engineering, 12(13s), 351-366.
[24] KC, A. S. P. (2024). Cloud-Based Top-Down and Bottom-Up Approach for Agriculture Data Integration. International Journal of Intelligent Systems and Applications in Engineering, 12(2), 346-354.
[25] Walsh, J., Mangina, E., & Negrão, S. Advancements in Imaging Sensors and AI for Plant Stress Detection: A Systematic Literature Review. Plant Phenomics.
[26] Munivel, M., & Enigo, V. F. (2024). MLIBT: A multi-level improvised binarization technique for Tamizhi inscriptions. Expert Systems with Applications, 236, 121320.
[27] Kasera, R. K., Gour, S., & Acharjee, T. (2024). A comprehensive survey on IoT and AI based applications in different pre-harvest, during-harvest and post-harvest activities of smart agriculture. Computers and Electronics in Agriculture, 216, 108522.
[28] Sivasubramaniyan, Thylashri, and Rajalakshmi Nagarnaidu RajaPerumal. "Original Research Article Identifying land use land cover dynamics using machine learning method and GIS approach in Karaivetti, Tamil Nadu." Journal of Autonomous Intelligence 7, no. 3 (2024).
[29] Saha, A., & Pal, S. C. (2024). Modelling groundwater vulnerability in a vulnerable deltaic coastal region of Sundarban Biosphere Reserve, India. Environmental Geochemistry and Health, 46(1), 8.
[30] Gunasekaran, S. (2024). Caste, food and colonialism:‘Outcaste’domestics in the European houses of Madras Presidency. The Indian Economic and Social History Review, 00194646231222641.
[31] Kizhakkeveettil, A., Parla, J., Patwardhan, K., Sharma, A., & Sharma, S. (2024). History, Present and Prospect of Ayurveda. In History, Present and Prospect of World Traditional Medicine (pp. 1-72).
[32] Akhter, M., Ullah, M. R., Khan, A. B. S., Hasan, M. M., Yasmin, F., Bosu, A., ... & Mahmud, Y. (2024). Assessment of different solvents effect on bioactive compounds, antioxidant activity and nutritional formation of red seaweed, Gracilariopsis longissima, from Bay of Bengal, Bangladesh. Discover Food, 4(1), 1-12.
[33] Gopalakrishnan, A., Sidden, C., Magudeeswaran, S., Bangaru, S., & Ramasamy, A. (2024). Synthesis, Crystal Structure, Spectroscopic and Computational Studies of NLO Active 2-Methylimdazolium 4-Nitrobenzoic Acid Single Crystal. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 123936.