JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

CROP YEILD PREDICTION AND RECOMMENDATION SYSTEM USING MACHINE LEARNING

Ketki R. Ingole, Kalyani D. Amalkar, Megha U. Kagliwal

Assistant Professor, Student, Student
Computer Science and Engineering
SIPNA College of Engineering and Technology, Amravati, India

Abstract: The Crop Yield Prediction System utilizes advanced technologies and machine learning algorithms to analyze historical data on crop yields, weather, soil quality, and agronomic practices. It provides accurate predictions of future crop yields, accessible through user-friendly web and mobile applications. Customization features enable farmers to tailor predictions to their specific needs, while a feedback mechanism continuously refines the models for improved accuracy. Real-time weather data integration enhances prediction accuracy, and decision support tools offer recommendations for optimal planting times and agronomic practices. Similarly, the Crop Price Recommendation System employs machine learning and real-time data to provide accurate crop price estimations, aiding farmers and stakeholders in making informed decisions for enhanced profitability and sustainability in agriculture.

Index Terms **- Crop Yield Prediction System, Machine learning algorithms, Historical data analysis, Weather data, Decision support tools, Crop Price Recommendation System**

I. Introduction

The modern agricultural landscape is experiencing a transformative evolution driven by the imperative to sustainably feed a growing global population amidst the challenges of climate change and resource constraints. Amidst this backdrop, the introduction of the Crop Yield Prediction System stands as a beacon of innovation, integrating cutting-edge technologies, notably machine learning and data analytics, to tackle the complex task of forecasting crop yields. This system promises to redefine how farmers and stakeholders navigate contemporary agriculture, offering a predictive framework poised to enhance productivity, sustainability, and informed decision-making.

At its core, the Crop Yield Prediction System employs a sophisticated data integration process, amalgamating diverse datasets encompassing historical crop yields, weather dynamics, soil quality parameters, and agronomic practices. Machine learning algorithms play a pivotal role, discerning intricate patterns within this data to forecast future crop yields accurately. The system's user-friendly interface, accessible through web and mobile applications, ensures accessibility to users of varying technological proficiency. Customization features enable tailoring predictions to specific crop types, geographical locations, and farming practices, fostering adaptability across diverse agricultural landscapes.

Moreover, the system incorporates a feedback mechanism to refine machine learning models continually, leveraging user input and additional data for ongoing improvement. Real-time integration of weather and climate data further enhances prediction accuracy, considering the dynamic impact of changing weather conditions on crop yields. Beyond mere prediction, the system offers decision support tools empowering farmers with actionable recommendations, from optimal planting times to agronomic practices. Scalable across various agricultural scales, the Crop Yield Prediction System aligns with precision agriculture and sustainable farming practices, standing as a transformative force in global food security and agricultural sustainability.

Transitioning to agricultural economics, the Crop Price Recommendation System aims to revolutionize crop pricing through data-driven technologies such as machine learning and big data analytics. By considering a wide array of factors influencing crop prices, including historical yield data, market trends, and government policies, this innovative system generates accurate and timely crop price estimations. Adaptive to changing market conditions, it assists stakeholders in making strategic pricing decisions, optimizing pricing strategies, and fostering fair and sustainable pricing practices.

This data-driven approach has the potential to revolutionize agricultural economics, fostering fair pricing, and sustainable growth within the agriculture sector. Through an exploration of factors influencing crop yield reconstruction, methodologies for crop price estimation, and potential applications, this seminar report delves into the transformative impact of the Crop Price Recommendation System on agricultural economics, paving the way for a more resilient and prosperous agriculture industry.

II. Literature Review

The literature surrounding crop yield prediction and recommendation systems reveals a rapidly advancing field that holds tremendous promise for revolutionizing agricultural practices. Reddy and Kumar (2021) showcased the increasing utilization of machine learning algorithms in predicting crop yields, emphasizing their potential to enhance accuracy and efficiency in agricultural decision-making. Similarly, the comprehensive evaluation by Thampi et al. (2020) sheds light on the complex relationship between climate parameters and crop performance, crucial for understanding the effects of changing climates on agricultural outcomes.

Pande et al. (2021) introduced the concept of a Crop Recommender System, highlighting the integration of machine learning for precise crop recommendations. Understanding and adapting to changing climates emerge as recurring themes, as illustrated in Yadav et al.'s (2022) study on the impacts of heat stress on wheat. In addition to considering crop yield and market demand, some researchers have emphasized theimportance of nutritional requirements in crop recommendations. Analyzing the nutrient content of different crops and their contribution to a balanced diet can help optimize food security and address nutritional deficiencies in specific regions (Girish et al., 2018).

Thapaswini and Gunasekaran's (2022) work shifts the focus to economic considerations, proposing a methodology for predicting crop prices using machine learning. This interdisciplinary approach underscores the significance of considering economic factors alongside agricultural practices. Additionally, Bhansali et al. (2022) present a system addressing both crop prediction and disease detection, emphasizing the holistic nature of modern agricultural technologies.

Sarosi and Anbarasi (2021) explore crop yield prediction using multi-parametric deep neural networks, highlighting the evolving landscape of predictive analytics in agriculture. Similarly, Flavarawan and Vincent (2022) advocate for the integration of deep reinforcement learning to address sustainability concerns in agriculture, showcasing innovative approaches to optimizing resource use.

The systematic literature review conducted by Thomassen-Klompenhouwer et al. (2020) provides a comprehensive overview of existing studies in crop yield prediction using machine learning, emphasizing advancements, challenges, and emerging trends in the field. This review serves as a valuable resource for understanding the current state of research and guiding future directions in agricultural data analytics.

Kursa et al. [19] implemented Boruta, an all-relevant feature selection method which gathers every feature that is critical to the outcome in certain circumstances. By contrast, most traditional feature selection algorithms follow a minimally optimal method in which they rely on a small subset of features that yield a minimal error on a selected classifier. Marcano Cedeno et al. [20], proposed a feature selection method based on sequential forward selection and the feed forward neural network to find the prediction error as a criterion for selection.

Zahra Karimi et al. [21], implemented a feature ranking method using a hybrid filter feature selection scheme for intrusion detection in a standard dataset. The experimental results show that the proposed technique offers higher accuracy than other methods. Surabhi Chouhan et al. [22], proposed a hybrid combination method of applying the Particle Swarm Optimization – Support Vector Machine (PSO-SVM) to select features from a dataset. Assorted benchmark datasets were tested with this technique.

David Heckmann et al. [23] described that the harnessing of natural variability in photosynthesizing ability as a way to improve yields, through a functional phylogenetic analysis for large-scale genetic screening is a laborious task. The potential for leaf reflectance spectroscopy to estimate photosynthetic efficiency specifications in Brassica oleracea and Zea mays, a C3 and a C4 seed, respectively, were analyzed, the findings show that phenotyping leaf reflectance is an effective method to enhance the photosynthetic ability of crops.

Maya Gopal and Bhargavi [24], proposed a wrapper feature selection method featuring Boruta that extracts features from a dataset for crop prediction. The technique improves prediction performance and provides effective predictors. In Boruta, the Z score has the most accurate measure, since it takes into consideration the variability of the mean loss of accuracy among trees in a forest.

In summary, the integration of machine learning, deep learning, and other advanced technologies holds immense potential for optimizing agricultural practices in response to evolving climatic conditions and economic considerations. By leveraging interdisciplinary approaches and innovative methodologies, researchers are paving the way for a more resilient and sustainable agricultural future.

III. Problem Statement

The agriculture sector faces numerous challenges, including climate variability, resource constraints, market fluctuations, and pest/disease outbreaks, which significantly impact crop production and profitability. Farmers require accurate and timely information to make informed decisions regarding crop selection, planting schedules, resource allocation, and pricing strategies. However, traditional methods of predicting crop yields and recommending crops often lack precision and fail to consider dynamic environmental factors, leading to suboptimal outcomes and economic losses.

There is a critical need for an advanced Crop Prediction and Recommendation System that integrates cutting-edge technologies, such as data analytics, machine learning, and real-time data processing, to address the complexities of modern agriculture. This system must effectively analyze a wide range of parameters, including historical crop yields, weather data, soil characteristics, market trends, and pest/disease incidence, to generate accurate predictions and recommendations tailored to specific locations and farming practices.

Key challenges to be addressed by the Crop Prediction and Recommendation System include:

1. Prediction Accuracy: Existing methods of crop yield prediction often lack accuracy, leading to unreliable forecasts and suboptimal decision-making by farmers. The system must leverage advanced data analytics and machine learning algorithms to improve prediction accuracy and provide reliable forecasts of future crop yields.

2. Adaptability to Climate Variability: Climate change is causing increased variability in weather patterns, posing challenges for crop production. The system must incorporate climate change scenarios and future weather projections to anticipate and adapt to changing environmental conditions, ensuring resilient crop recommendations.

3. Integration of Diverse Data Sources: Agricultural data is diverse and often fragmented, including information from multiple sources such as satellite imagery, weather stations, soil databases, market reports, and farmer surveys. The system must seamlessly integrate these data sources to provide comprehensive insights and recommendations.

4. Real-Time Data Processing: Timely access to real-time data is essential for effective decision-making in agriculture. The system must employ efficient data processing techniques and cloud computing technologies to analyze large volumes of data in real-time, enabling farmers to make timely adjustments to their farming practices.

5. User-Friendly Interface: Farmers may have varying levels of technical expertise and access to technology. The system must have a user-friendly interface, accessible through web and mobile applications, that provides clear and actionable recommendations to farmers in a simple and intuitive manner.

6. Feedback Mechanism: Continuous feedback from farmers is essential for refining prediction models and improving recommendation accuracy over time. The system must incorporate a feedback mechanism that allows farmers to provide input on the performance of predictions and recommendations, enabling iterative improvement of the system.

7. Scalability and Accessibility: The system must be scalable to accommodate a large number of users and accessible to farmers in diverse geographical locations, including those in remote and rural areas with limited internet connectivity.

Overall, the development of an advanced Crop Prediction and Recommendation System presents an opportunity to revolutionize agricultural decision-making, empower farmers with valuable insights, and enhance the productivity and sustainability of the agriculture sector. By addressing the challenges outlined above, the system can contribute to improved crop yields, reduced input costs, increased profitability, and food security for farming communities worldwide.

IV. PROBLEM ANALYSIS

Climate change has led to increased variability in weather patterns, including fluctuations in temperature, rainfall, and extreme weather events. These variations can significantly impact crop growth, pest and disease prevalence, and overall agricultural productivity. Farmers need accurate predictions and recommendations to adapt their practices to changing climatic conditions and minimize risks associated with climate variability. Traditional methods of predicting crop yields often rely on historical data and simplistic models, leading to limited accuracy and reliability. Variability in environmental factors, soil conditions, pest infestations, and market dynamics introduces uncertainty into yield predictions. Farmers require more sophisticated prediction models that incorporate a wide range of variables and leverage advanced analytical techniques to improve accuracy and provide reliable forecasts.

Agricultural data is often fragmented and dispersed across various sources, including government agencies, research institutions, and private companies. Integrating these diverse datasets and making them accessible to farmers can be challenging due to technical barriers, data privacy concerns, and limited infrastructure in rural areas. Farmers need a centralized platform that aggregates and analyzes relevant data from multiple sources to provide comprehensive insights and recommendations. Crop prices are influenced by various factors, including supply and demand dynamics, global market trends, currency exchange rates, and trade policies. Market volatility and price fluctuations can affect farmers' profitability and financial stability. Farmers require accurate predictions of crop prices and recommendations for pricing strategies to maximize their revenue and mitigate risks associated with market uncertainty.

Soil health plays a crucial role in determining crop productivity and nutrient uptake. Soil characteristics, such as pH levels, nutrient content, organic matter, and texture, influence crop suitability and yield potential. Understanding soil health and matching crops to specific soil requirements can optimize yields and minimize the need for chemical inputs. Farmers need recommendations for crop selection and soil management practices that promote soil health and sustainable agriculture. Pest infestations and disease outbreaks pose significant threats to crop production and food security. Monitoring and predicting pest and disease incidence require timely data on environmental conditions, pest life cycles, and crop susceptibility. Farmers need recommendations for integrated pest management strategies, including crop rotation, pest-resistant crop varieties, and biological control methods, to minimize crop losses and reduce reliance on chemical pesticides.

The adoption of technology and innovation in agriculture, such as precision farming techniques, remote sensing technologies, and digital agriculture tools, has the potential to improve productivity, efficiency, and sustainability. However, many farmers face challenges in accessing and adopting these technologies due to cost barriers, lack of training, and limited awareness of their benefits. Farmers need recommendations for adopting appropriate technologies and practices that align with their resources, capabilities, and objectives. Engaging farmers in the development and implementation of crop prediction and recommendation systems is crucial for ensuring relevance, usability, and acceptance. Farmers' feedback on the performance and usability of prediction models and recommendations is essential for continuous improvement and refinement. Establishing effective communication channels and feedback mechanisms between farmers, researchers, and technology developers can enhance user engagement and satisfaction.

By conducting a thorough problem analysis, stakeholders can identify the key challenges and opportunities in crop prediction and recommendation and develop targeted solutions to address farmers' needs and improve agricultural outcomes.

V. IMPLEMENTATION STRATEGY

The Data Processing stage in the Crop Price Recommendation System involves the comprehensive integration and analysis of various parameters to generate accurate and dynamic crop price recommendations. Leveraging cutting-edge data processing techniques and machine learning algorithms, the system ensures timely and informed pricing decisions. Key parameters considered in this stage include:

1. Global Commodity Prices: Real-time data on global commodity prices is integrated to understand international market trends, enabling the system to anticipate price changes and provide strategic recommendations.

2. Currency Exchange Rates: Fluctuations in currency exchange rates impact revenue from crop exports. The system considers these changes to provide accurate price estimations for countries involved in agricultural trade.

3. Transportation Costs: Incorporating fuel prices, distance to markets, and logistics expenses ensures fair pricing while accounting for the cost of transporting produce from farms to markets.

4. Supply Chain Information: Analysis of supply chain dynamics helps the system understand how intermediaries affect crop prices, providing insights into potential price variations.

5. Technological Advancements: Consideration of technological factors, such as precision agriculture tools and mechanization, assesses their impact on crop yields and pricing.

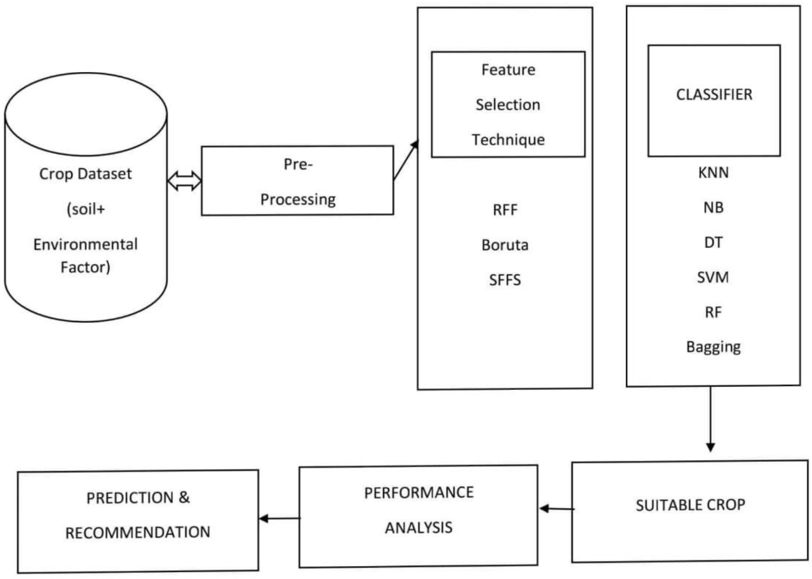
6. Soil Health: Data on soil pH levels, nutrient content, and texture are considered to match crops with specific soil requirements, optimizing yield potential.

7. Pest and Disease Incidence: Historical data on pest and disease incidence assesses the risk of crop damage, enabling the system to recommend pest-resistant crops to mitigate potential losses.

8. Market Demand and Trends: Analysis of market data and consumer preferences informs recommendations for crops with higher market value and demand.

By processing and integrating these parameters, the Crop Price Recommendation System generates accurate, region-specific crop price recommendations. This data-driven approach empowers stakeholders with valuable insights for setting competitive and profitable prices, ultimately improving agricultural practices and economic growth in the sector.

In implementing these solutions, the data-driven Crop Prediction System can overcome the limitations of traditional techniques. By providing farmers with informed decisions and recommendations based on historical and real-time data, they can optimize crop choices, increase yields, and contribute to sustainable and profitable farming practices. Key steps in implementing the Crop Prediction System include data collection, preprocessing, model development, recommendation engine creation, user interface design, feedback loop implementation, scalability, security, and integration with existing agricultural technology platforms.



Data Pre-processing Diagram for Prediction and Recommendation of Crops

VI. FLOWCHART

A flowchart of a crop prediction and recommendation system outlines the step-by-step process involved in generating predictions and recommendations for agricultural decision-making. Here's a brief overview of the components typically included in such a flowchart:

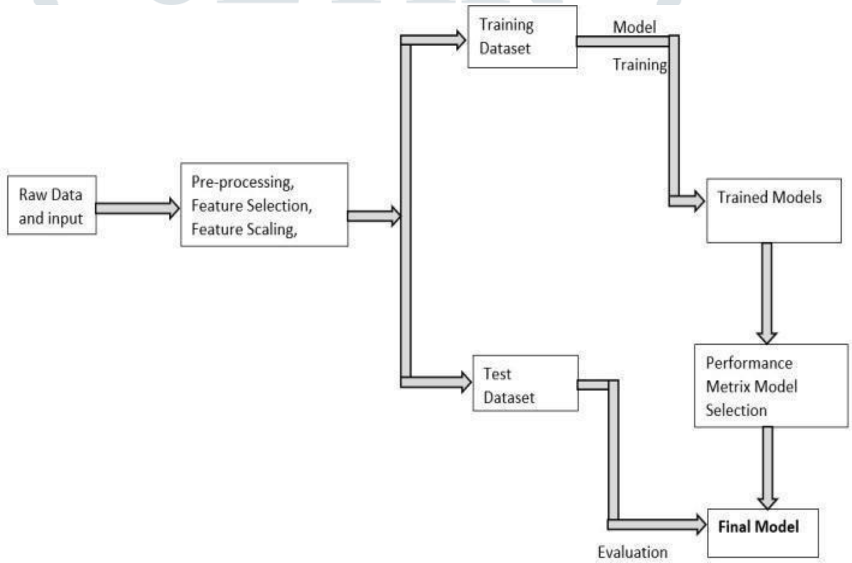
1. Data Collection: The process begins with gathering relevant data from various sources, including historical crop yields, weather patterns, soil characteristics, market prices, and pest/disease incidence. This data serves as the foundation for analysis and prediction.

2. Data Preprocessing: The collected data undergoes preprocessing to clean and transform it into a suitable format for analysis. This step involves handling missing values, removing outliers, and normalizing data to ensure consistency and accuracy.

3. Feature Selection and Engineering: Relevant features or variables are selected from the dataset, and new features may be engineered to enhance prediction accuracy. This step involves identifying the most informative predictors and creating composite variables that capture important relationships.

4. Model Training: Machine learning models are trained using the preprocessed data to learn patterns and relationships between input variables (e.g., weather, soil) and target variables (e.g., crop yield, price). Various algorithms such as regression, neural networks, or ensemble methods may be used for training.

5. Model Evaluation: The trained models are evaluated using validation datasets to assess their performance and generalization ability. Metrics such as accuracy, precision, recall, and F1-score are calculated to gauge the model's predictive power.



Flowchart for Machine Learning Model of Crop Recommendation System

VII. PROPOSED INITIATIVES AND GOALS

The proposed methodology of the research work begins with the preprocessing of input data, which involves tasks such as handling missing values, eliminating redundant data, standardizing the dataset, and converting target attributes into factor attributes. Following this, essential attributes are extracted using wrapper feature selection techniques. The optimized attributes undergo classification techniques after splitting the dataset into training and testing phases.

Unknown samples from the training dataset are utilized to train the classification algorithm to determine the most suitable crop for cultivation in a specific area of land. Subsequently, the testing dataset is employed to predict the crop to be raised using the trained classifier. Finally, the suitability of the selected crop is evaluated using various performance metrics, which helps in determining the best feature selection technique paired with an appropriate classification method.

Building upon past systems, the proposed system suggests crops based on soil classification by employing an ensemble of classifiers, namely Bagged Tree, Naive Bayes, and Support Vector Machine (SVM) algorithms. This amalgamation aims to enhance the accuracy of the system and provides a list of suitable crops according to the soil type.

In future iterations, advanced classification algorithms and techniques can be explored to further improve the system's accuracy across various datasets. Additionally, incorporating a location recommendation module based on crop suggestions could enhance the practicality of the system by recommending suitable locations for cultivating suggested crops.

The proposed model is designed to be computationally efficient, making it suitable for deployment on lightweight capability devices. However, it requires a larger number of parameters for better accuracy. Future improvements could focus on enhancing data collection methods, integrating emerging technologies like blockchain for transparent pricing data, and incorporating climate change scenarios to prepare for future challenges in agriculture. Additionally, efforts can be directed towards reducing the training time of the model and incorporating self-learning capabilities for continual improvement.

VIII. REFERENCES

1. S. Das. Filters, wrappers and a boosting-based hybrid for feature selection. In International conference on machine learning 1 (2001), pp. 74–81. [Google Scholar]

2. H. Liu and Y. Lei, toward integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng. 17 (2005), pp. 491–502. doi:https://doi.org/10.1109/TKDE.2005.66. [Crossref], [Web of Science ®], [Google Scholar]

3. A. Maruf, M. Abdullah, and S. Shatabda, iRSpot-SF: Prediction of recombination hotspots by incorporating sequence-based features into Chou's Pseudo components, Genomics 111 (2019), pp. 966–972. doi:https://doi.org/10.1016/j.ygeno.2018.06.003. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]

4. I. Kononenko, Estimating attributes: Analysis and extensions of RELIEF. European conference on machine learning 784 (1994), pp. 171–182. [Crossref], [Google Scholar]

5. J. Pavel Pudil, N.N. Choakjarernwanit, and J. Kittler, Feature selection based on the approximation of class densities by finite mixtures of special type, Pattern Recognit 28 (1995), pp. 1389–1398. doi:https://doi.org/10.1016/0031-3203(94)00009-B. [Crossref], [Web of Science ®], [Google Scholar]

6. J. Novovicová, P. Pudil, and J. Kittler, Divergence based feature selection for multimodal class densities, IEEE Trans. Pattern Anal. Mach. Intell. 18 (1996), pp. 218–223. doi:https://doi.org/10.1109/34.481557. [Crossref], [Web of Science ®], [Google Scholar]

7. H. Liu and H. Motoda, Feature Selection for Knowledge Discovery and Data Mining, 1st ed., Springer, US, 1998. [Crossref], [Google Scholar]

8. M. Dash, K. Choi, P. Scheuermann, and H. Liu. Feature selection for clustering-a filter solution. IEEE International Conference on Data Mining, 2002, Proceedings., Maebashi City, Japan, pp. 115–122. [Crossref], [Google Scholar]

9. I. Guyon, S. Gunn, M. Nikravesh, and L.A. Zadeh (eds.), Feature Extraction: Foundations and Applications, 1st ed., Springer-Verlag, Berlin Heidelberg, 2008. [Google Scholar]

10. I. Guyon and A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003), pp. 1157–1182. [Crossref], [Google Scholar]

11. W. Paja, K. Pancerz, and P. Grochowalski, Generational feature elimination and some other ranking feature selection methods, in Advances in Feature Selection for Data and Pattern Recognition, International Publishing, Springer, Charm, Vol. 138, 2018, pp. 97–112. [Crossref], [Google Scholar]

12. J.-Y. Hsieh, W. Huang, H.-T. Yang, C.-C. Lin, Y.-C. Fan, and H. Chen. Building the rice blast disease prediction model based on machine learning and neural networks., Easy chair the world of scientists (2019), pp. 1–8. [Google Scholar]

13. R. Kohavi and G.H. John, Wrappers for feature subset selection, Artif. Intell. 97 (1997), pp. 273–324. doi:https://doi.org/10.1016/S0004-3702(97)00043-X. [Crossref], [Web of Science ®], [Google Scholar]

14. I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, Gene selection for cancer classification using support vector machines, Mach Learn 46 (2002), pp. 389–422. doi:https://doi.org/10.1023/A:1012487302797. [Crossref], [Web of Science ®], [Google Scholar]

15. M. Sebban and R. Nock, A hybrid filter/wrapper approach of feature selection using information theory, Pattern Recognit 35 (2002), pp. 835–846. doi:https://doi.org/10.1016/S0031-3203(01)00084-X. [Crossref], [Web of Science ®], [Google Scholar]