WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Crop Yield Prediction Using Machine Learning

***Vishnu Chavan*[1] *, Mallamma V Reddy*[2]**

[1] MCA Student, Department of Computer Science Rani Channamma University, Belagavi
[2] Assistant Professor, Department of Computer Science Rani Channamma University, Belagavi

ABSTRAC T

Farming is the foundation of the Indian economy with over half of the population relying on it for their livelihood. Changes in weather, climate, and other environmental factors have emerged as a significant threat to the successful practice of agriculture. Machine learning is crucial for Crop Yield Prediction (CYP) by serving as a decision support tool for determining which crops to plant and how to manage them throughout the growing season. The current study focuses on a systematic review that identifies and combines the characteristics through artificial intelligence approaches. The primary constraints of the Neural Network are a decrease in the relative error and reduced efficiency in predicting Crop Yield. Likewise, supervised learning methods struggled to understand the complex relationship between input and output variables when it came to the grading or sorting of fruits. Numerous studies were suggested to improve agriculture, focusing on developing a precise and efficient model for classifying crops like estimating crop yield using factors such as weather, crop disease, and growth phase. This paper investigates different machine learning methods used in crop yield RANI CHANNAMMA UNIVERSITY BELAGAVI PAGE 1 employed for CYP, along with multiple methods devised for analyzing crop yield prediction estimation and analyzes their accuracy.

KEYWORDS: Machine learning, processing agriculture, soil analysis, climate condition, data-driven,crop yield prediction.

1. INTRODUCTION

In terms of farm production, India holds the second position globally. In 2009, agriculture and related sectors, such as forestry and fishing, contributed 16% to the GDP and employed around half of the population. The financial impact of agriculture on India's GDP is decreasing consistently. Crop yields depend on various factors such as climate, geography, organic practices, political factors, and financial circumstances. Farmers face challenges when deciding which crops to grow, especially when market prices are uncertain. According to Wikipedia statistics, the suicide rate among farmers in India ranged between 1.4 and 1.8 per 100,000 total population over a 10-year period until 2005. However, there was a significant increase in farmer suicides with 5650 reported in 2014, which escalated to over 8000 in 2015. nowadays raising awareness about agriculture using technology has become a must food insecurity is a result of seasonal climate changes that also threaten basic resources like soil water and air in a situation where crop yield rates are regularly not keeping up with demand a smart system that can address the issue of declining crop yield is required consequently in order to solve this issue we suggest choose their crops based on environmental and economic considerations maximizing their yield and eventually assisting in meeting the nation's growing need for food supply machine learning is used by the suggested system to generate the predictions in order to maximize crop output for farmers the system will provide crop yield and crop selection depending on meteorological attributes suitable for the crop the technology forecasts agricultural yields by examining variables including precipitation temperature area in hectares season etc the system also helps in suggesting whether a particular time is the right one to use fertilizers. Crop prediction is a significant issue in agriculture. Each farmer consistently aims to ascertain the quantity of yield that will be generated and if it meets their anticipated level. Previously, the prediction of yield was based on analyzing farmer's prior experience with a specific crop. The crop harvest is mainly influenced by weather, pests, and the planning of when to harvest. Precise data on crop yield history is crucial for informing decisions regarding agricultural risk management.

2. LITERATURE SURVEY:

Crop prediction is a critical aspect of modern agriculture, leveraging machine learning and data analytics to enhance decision-making for farmers. Recent studies highlight the integration of various machine learning algorithms, such as Decision Trees, Random Forest, and Support Vector Machines, to predict crop types based on soil and environmental features, achieving high accuracy rates of up to 98%[2]. Additionally, weather data, including temperature and precipitation, plays a significant role in forecasting crop yields, with advanced techniques like remote sensing further refining predictions[4]. The development of interactive prediction systems empowers farmers to make informed choices regarding crop selection and management practices[3]. However, challenges such as data complexity and the need for effective feature selection must be addressed to optimize model performance[5]. Overall, the combination of these technologies promises to improve agricultural productivity and sustainability in the face of changing environmental conditions[1].

3. PROPOSED METHODOLOGY:

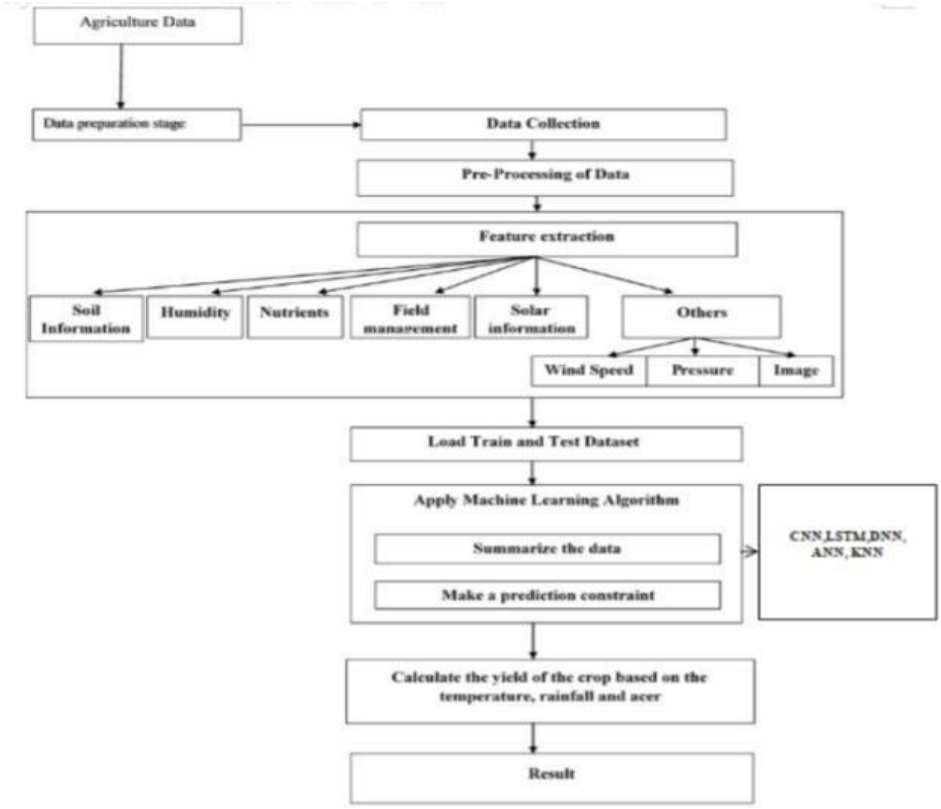


Figure 1. System Architecture

The majority of the existing models used neural networks, random forests, and KNN regression methods for predicting CYP, while a range of machine learning techniques were also employed for optimal predictions. Data collection: Compile a variety of voice samples from people of various genders. To accommodate for variances and guarantee inclusion, the dataset should include a broad variety of voices, dialects, languages, and speech patterns.

1. Creation, repair and maintenance of ML algorithms required huge costs as they are very complex.
2. ML technique used for Crop yield prediction (mustard, wheat) combined input and output data but failed to obtain better results statistically
3. Due to the nature of linear connection in the parameters, the regression model was failed to provide the exact prediction in a complex situation such as extreme value data and nonlinear data.
4. The current K-NN models were employed for yield prediction classification, but their performance was reduced by the nonlinear and highly adaptable issues found in KNN.

They were operated in a locality model that incremented the dimensionality of the input vector made confusion for classification.

1. An appropriate decision was not taken during classification because a fewer quantity of data was available for estimation of crop yield.

Exploratory Data Analysis

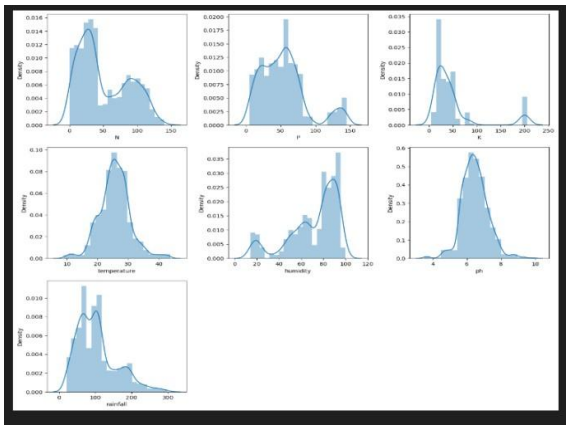


Fig : 2 Data Analysis

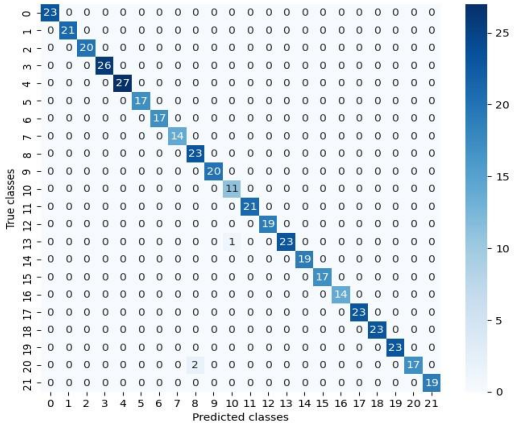


Fig: 3 Predicted class

Algorithms Used

There are many different algorithms available in machine learning. The categories for these are often association, grouping, regression, and classification. Clustering and association fall under unsupervised learning, but classification and regression methods go under supervised learning

Table 1.1 Summary of the Approaches

Sno.	Algorithm	Accuracy
1.	Support Vector Machine (SVM)	75%
2.	Decision Tree	60%
3.	Random Forest	95%
4.	K Nearest Neighbours (KNN)	90%

- SVM vs. Random Forest: The best way to be worry-free is probably to use random forests. The number of trees—generally speaking, the more the better—may be the only true hyperparameter to adjust. On the contrary, there are a lot of knobs to be turned with SVMs: Choosing the ―right‖ kernel, regularization penalties, the slack variable, etc. Random forests are significantly easier to find a decent, reliable model for and much easy to train. In addition to the usual reasonable amount of parameter tuning required for SVMs, the computing cost also increases linearly with the number of classes.
- Deccision vs. Random Forest: Linear regression is rarely as effective as random forests. Without transforming the data, random forests provide a superior first fit. In practically every aspect, they are more tolerant. Scaling your data or doing any monotonic transformations (log, etc.) are not necessary. Frequently, eliminating outliers is not even necessary. If you provide categorical categories, the data will be automatically partitioned to improve the fit.
- KNN vs. Random Forest: KNN's costly real-time execution makes Random Forest quicker. While there is no such choice to be made in Random Forest, _K' should be chosen carefully. If the sample size is huge, KNN has a high calculation cost during runtime.

Random Forest

Even without hyper-parameter tweaking, Random Forest is a versatile and user- friendly machine learning technique that typically yields excellent results. Due to its versatility and simplicity, it is also one of the most used algorithms. It can be applied to challenges involving regression as well as classification. For a more precise and reliable forecast, Random Forest constructs several decision trees and combines them. Random forest has the main benefit of being applicable to both classification and regression problems, which make up the bulk of existing machine learning systems. The random forest algorithm's ease of measuring the relative contribution of each feature to the prediction is another excellent feature. Sklearn is a fantastic resource for score automatically for each feature after training and scales the results so all importance is equal to one.

Random Forest operates in the following ways:

Step 1 − Start by choosing random samples from the provided dataset.

Step 2 − For each sample, this algorithm will then create a decision tree. Next, it will obtain each decision tree's Prediction result.

Step 3 − Voting will be done for each anticipated outcome in this step.

Step 4 − Decide which prediction result received the most votes to determine the ultimate outcome.

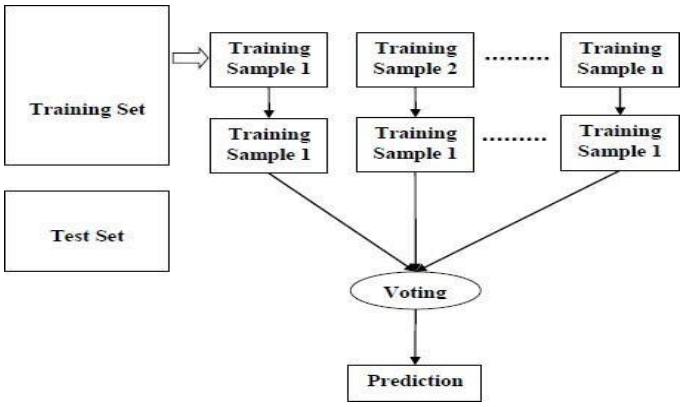


Fig : 4 Random Forest Flow

MODULES:

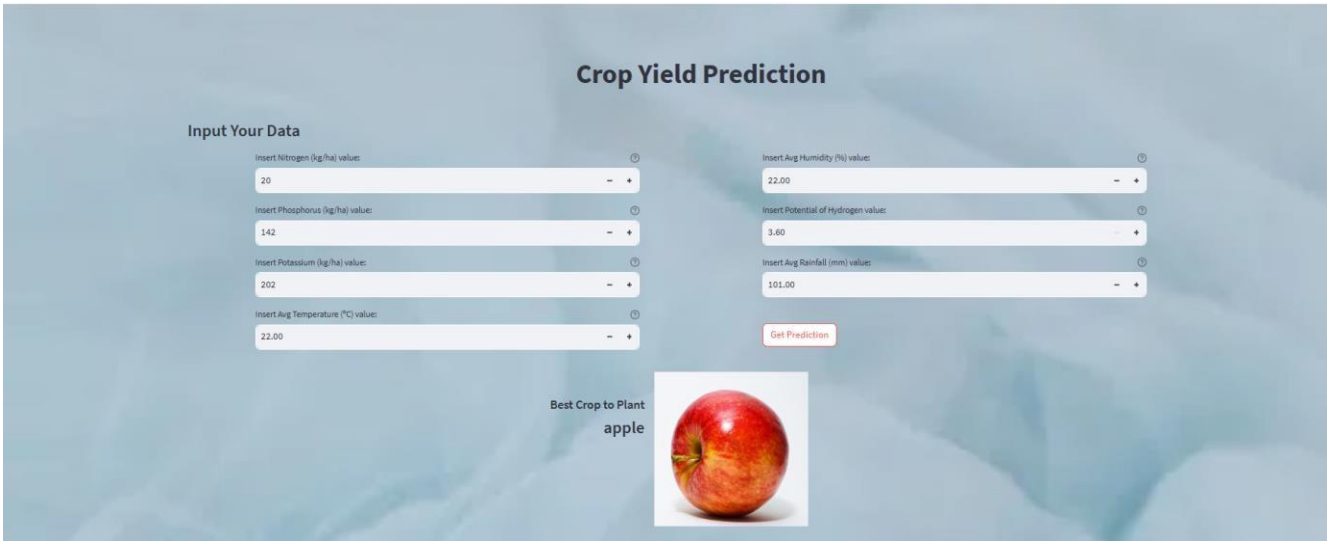
Upload Crop Dataset: The dataset for crop production, which predicts crop name and yield, is inputted into classification and regression algorithms.

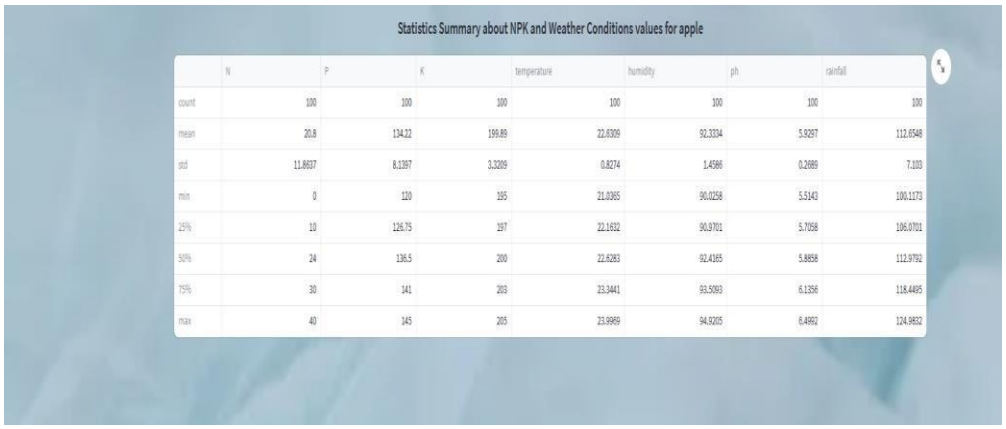
Preprocess Dataset: Research was carried out using the dataset from the Indian government, showing that the Random Forest Regressor provides the most accurate predictions for yield. Simple Recurrent Neural Network outperforms in predicting rainfall, while LSTM is more effective in predicting temperature. Yield prediction for a specific district can be achieved by considering factors such as rainfall, temperature, season, and location.

Train Machine Learning: This pertains to predicting the yield in each district based on the crop being cultivated in that district. Yield projections are being made for different crops in each district and identifying which crops have the highest yields.

Upload Test Data &Predict Yield: Findings indicate that Random Forest outperforms all other classifiers when all parameters are considered together. This will assist farmers in selecting the appropriate crop for the upcoming season and also close the technology and agriculture sector divide.

RESULTS





	N	P	K	temperature	humidity	ph	rainfall
count	100	100	100	100	100	100	100
mean	20.8	134.22	199.89	22.6309	92.3334	5.9297	112.6548
std	11.8637	8.1397	3.3209	0.8274	1.4586	0.2689	7.103
min	0	120	195	21.0365	90.0258	5.5143	100.1173
25%	10	126.75	197	22.1632	90.9701	5.7058	106.0701
50%	24	136.5	200	22.6283	92.4165	5.8858	112.9792
75%	30	141	203	23.3441	93.5093	6.1356	118.4495
max	40	145	205	23.9969	94.9205	6.4992	124.9852

Fig 5: Prediction

TESTING:

Software testing is an inquiry carried out to notify stakeholders about the caliber of the software product or service being tested. Software testing can also offer an impartial, unbiased perspective on the program, enabling the company to recognize and comprehend the risks associated with using the program. In order to detect software bugs (errors or other problems) and confirm that the software product is suitable for use, test approaches involve running a program or application.

In software testing, a system or software component is executed in order to assess one or more properties of interest. These characteristics typically show how much the part or system being tested:

- meets the conditions that guided its design and development**,**
- responds rightly to all kinds of inputs,
- performs its functions within an respectable time, can be installed and run in its intended surroundings
- Achieves the general affect its stakeholder ‘s desire.

Functionality Testing:

- The database connection has been made successfully.
- All the application's forms function as they should.

  When incorrect inputs are entered, appropriate alert messages are shown.

Following each operation performed on the application, the relevant data is retrieved from the database.

Usability Testing:

- Because the user provides input via a dropdown menu, the system receives accurate input.
- Incorrect inputs provided by the system are efficiently managed.
- The application's content is authentic and sourced from reliable sources.
- Accurate and well-balanced datasets were used to train the crop yield prediction models.

Interface Testing:

- The operation connects rightly with the garçon. In case of failure an applicable communication is displayed.
- Interruptions by the garçon or by the stoner are handled efficiently.
- If the stoner enters wrong credentials or invalid dispatch id, the operation handles it
- efficiently by displaying applicable dispatches.

Performance Testing:

- The connection is encrypted, and user information is maintained in a secure manner,it functions well with moderate internet speeds.
- User input is accurately logged, and responses are promptly captured.

HARDWARE AND SOFTWARE REQUIREMENTS:

Software Requirements:

1) Python
2) Djnago/Flask
3) Anaconda

Hardware Requirements:

1) Min 4GB RAM
2) Min 2.2GHZ or Higher Processor
3) Min 50Gb Free Space
4) Windows 8 or Higher

CONCLUSION

The goal of this method is to assist farmers in becoming more financially stable while also addressing the rising number of farmer suicides.

The Crop Recommender system assists farmers in selecting which crop to plant and in estimating the output of a particular crop. Additionally, it provides the user with the ideal timing for applying the fertilizer Machine learning techniques were used to gather, examine, and train relevant datasets. The system keeps track of the user's position and uses that information to retrieve necessary data from the backend. As a result, just basic information from the user—such as the region and kind of soil—is required. Agriculture benefits from the use of this system. The system's ability to forecast the weather for the next 14 days allows it to make one of its most significant and innovative contributions: it suggests to the user when to apply fertilizer. A list of crops and their productions according to climate conditions is also provided by the system.

FUTURE WORK

Future research will concentrate on determining the crop sequence that should be planted based on the soil and weather, as well as on periodically updating the databases to generate precise forecasts. A fully automated system that accomplishes the same is the goal of the Future Work. Our aim is to incorporate the feature of supplying the appropriate fertilizer for a certain crop and area. It is necessary to study fertilizers and how they interact with soil and climate in order to put this into practice. Our goal is also to anticipate crisis situations before they arise, such as the recent increase in onion prices.

ACKNOWLEDGEMENT

I am grateful to Dr.Mallamma V. Reddy Assistant Professor Department of Computer Science, Rani Channamma University, Belagavi for his valuable guidance for completion of this work

REFERENCES:

1) Vivian, Rawade., NULL, AUTHOR_ID. (2024). 1. Crop Prediction by Implementing Machine Learning in an IoT-Based System. doi: 10.1002/9781394242566.ch6
2) S, Nifasath, Piyar., S., Baulkani. (2024). 2. Crop Prediction based on Soil and Environmental Features Using Classification Models. doi: 10.1109/icdsis61070.2024.10594482
3) P., Srinivas, Karthik., Bolloju, Sanjith., Betha, Charan, Satya, Raj., G., Venkata, Ramana, Reddy., Bhavani, Vasantha. (2024). 3. Crop Yield Prediction. doi: 10.1109/conit61985.2024.10626796
4) Mr., V., Shanmugam., I., Sriteja., K., Sai, Dathu., K., Raju., S., Sai, Kumar., G., Karun. (2024). 4. Crop Yield Prediction using Machine Learning. International Journal of Advanced Research in Science, Communication and Technology, doi: 10.48175/ijarsct-18185
5) RADHA, KUMARI, C., Benson, Raja, B., Naveen, Kumar, S., Siva, kumar, S.., R., Vinoth. (2024). 5. A Smart Agriculture Crop Prediction System Using Various Feature Selection Techniques and Classifiers. International Journal of Research Publication and Reviews, doi: 10.55248/gengpi.5.0524.1343
6) Pamukuntla, Mahesh., Rajkumar, Soundrapandiyan. (2024). 7. Yield prediction for crops by gradient-based algorithms. PLOS ONE, doi: 10.1371/journal.pone.0291928
7) Muhammad, Umar, Abdullahi., Morufu, Olalere., Gilbert, I.O., Aimufua., Kene, Tochukwu, Anyachebelu., Bako, Halilu, Egga. (2024). 8. Crop Recommendation Predictive Analysis using Ensembling Techniques. Deleted Journal, doi: 10.33003/jobasr-2024-v2i1-43
8) Swathi., A. (2024). 9. Crop prediction. Indian Scientific Journal Of Research In Engineering And Management, doi: 10.55041/ijsrem28605
9) Rawatmal, B., Surana., Ritu, Khandelwal. (2024). 10. Crop Yield Prediction Using Machine Learning: A Pragmatic Approach. doi: 10.21203/rs.3.rs-4575893/v1
10) Groundnut Crop Yield Prediction Using Machine Learning Techniques by Vinita Shah and Prachi Shah.
11) Developing regression model to forecast the rice yield at Raipur condition by A Jain, JL Chaudhary, MK Beck and Love kumar.
12) K. A. Shastry, and H. A. Sanjay, "Hybrid prediction strategy to predict agricultural information," Applied Soft Computing, vol. 98, pp. 106811, 2021.
13) D. A. Bondre, and S. Mahagaonkar, "Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms," International Journal of Engineering Applied Sciences and Technology, vol. 4, no. 5, pp. 371-376, 2019.
14) B. Devika, and B. Ananthi, "Analysis of crop yield prediction using data mining technique to predict annual yield of major crops," International Research Journal of Engineering and Technology, vol. 5, no.12, pp. 1460-1465, 2018.
15) P. Sivanandhini, and J. Prakash, "Crop Yield Prediction Analysis using Feed Forward and Recurrent Neural Network," International Journal of Innovative Science and Research Technology, vol. 5, no. 5, pp. 1092-1096, 2020.
16) N. Nandhini, and J. G. Shankar, "Prediction of crop growth using machine learning based on seed," Ictact journal on soft computing, vol. 11, no. 01, 2020
17) P. Kumari, S. Rathore, A. Kalamkar, and T. Kambale, "Predicition of Crop Yeild Using SVM Approch with the Facility of EMART System"Easychair 2020.
18) S. D. Kumar, S. Esakkirajan, S. Bama, and B. Keerthiveena, "A microcontroller based machine vision approach for tomato grading and sorting using SVM classifier," Microprocessors and Microsystems, vol. 76, pp.103090, 2020