International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

An Efficient Ancient Tamil Script Classification System using Gradient Boosted Tree Algorithm

T.S.Suganya , S.Murugavalli

Abstract*: *A variety of historical information's has been derived and obtained from the various forms of inscriptions existing worldwide. Regional languages of certain selected sources have been employed in the inscription writing procedures. One such regional language with a wealthy literature and heritage that can be suitably opted for the inscription writing process is 'Tamil'. Encryption of the writing was done with the help of certain materials such as metals, conch shells, stones, palm leaf, and copper plate. These are found to be rich in the data contents related to the fields of history, religious, administrative, astronomy, culture, economic tax and educational conditions. This paper incorporates certain feature extraction methods like Shape and Hough transform and Gradient Boosting tree as classifiers for identifying the ancient Tamil script.

Keywords* : *About four key words or phrases in alphabetical order, separated by commas.

I. INTRODUCTION

The chronicled places in India have been found to possess huge portions of the engravings. These engravings must be protected and interpreted in appropriate ways. The archaeological department essentially performs the above mentioned tasks. The composition of these engravings can be done only on a certain selected surfaces such as the stones, metals, rocks, copper plate just as on palm leaf. Together with the selection of the medium meant for writing the most difficult job for the epigraphists is the selection of the graphemes, this appears to be tedious as their style appears to be huge, it possesses engravings with scratches, it these scratches that indicate the maturity level [1]. The duty of the Archaeological department is to completely safeguard these engravings, for which various actions have been encountered by the department, further they also involve themselves in creating the epigraphy. Computers are used for the process of digitising the engravings, for removing the unsettling influences similar to that of the breakages. For accomplishing the above tasks the computers are found to incorporate and utilize the various picture handling mechanisms. Other than safeguarding the epigraphy it is also mandatory to discover epigraphers who can peruse them.

During the ancient time frame, Tamil Nadu had been observed to possess lower Palaeolithic settlements. Various estimations had been performed for the determination of the existence, the results of these estimations proved that the settlements had existed from around 1,510,000 BCE until around 3000 BCE. A major duration of the Palaeolithic stage staged the human settlements near stream valleys where there exited a slight timberland spread meadow environment. South India comprises of only two territories of lower Palaeolithic progress till date, hence only a minimum number of occupants have been witnessed in these territories. The Archaeological department had uncovered the fossils of creatures just as crude stone apparatuses close to the northern Tamil Nadu district. These discoveries were perfectly approved and accepted to be a piece of 3,000,000 BCE. The populace in Southern India comprised of a variety of animal species, Homo erectus, and developed due in Palaeolithic 'old stone age' for a considerable period of time. It was also found that certain barely unprocessed actualizes such as hand axes and choppers were adopted and used by them to survive as hunter gatherers.

The Neolithic period existed in Tamil Nadu in the period of around 2500 BCE. It was during this Neolithic period that the various forms of stone instruments were pulverized and cleansed for the purpose of converting them into nice and beautiful equipments. In Tamil Nadu a Neolithic period based cleaver head possessing a piece of ancient writing was identified. Individuals surviving in the Neolithic period had opted for settlements on little level slants or on lower regions. The individuals in this period usually possessed temporary form of settlements as they had to move often from one place to another in search of brushing lands. These people followed the strategy of burying the dead inside large containers or inside the trenches. Further these people made use of the copper metal for making equipments and weapons. During the Iron Age people made use of the iron metal for making equipments and weapons. Numerous spots possessing the Megalithic entombment locales have been found to exist in the peninsular India suitably representing the culture of the Iron Age. Information's obtained from the unearthed materials just as that of the study and interpretation of the entombment landmarks has confirmed the movement of the Iron Age from the northern locales to the various districts in the southern locales.

Manuscript published on 30 September 2019
* Correspondence Author
T.S.Suganya*, Research Scholar, Department of Computer Science, Sathyabama Institute of Science and Technology, Chennai, Tamilnadu, India
Email: tssuganya@yahoo.com
Dr. S.Murugavalli, Professor & Head , Department Computer Science & Engineering, Panimalar Engineering College , Chennai,Tamilnadu, India
Email: murugavalli26@rediffmail.com

Unearthing procedures similar to that of the ones performed in the Northern zones were also performed in the areas of Adichanallur in the district of Thirunelveli in Tamil Nadu; the results obtained from these procedures proved that the Megalithic culture had moved towards the southern locales. It was the various writing works that suggested to the adoption of the funerary and the interment practices in the sangam period hundreds of years ago.

Among the various writings, Manimekalai (fifth century A.D) appears to be a famous and unique Buddhist epic suggested in section 6 (66-67) refers to the various types of committal procedures, post defleshing interment, casing the expired in a pit, shake chamber or cist entombment , urn interment uncapped with top are few of them. The above mentioned entombment methodologies were found to exist even during the Sangam age possessing autonomy in a society that was in its developing stage. Sanctuaries have been developed with the locally attainable stones and this strategy has taken a form by deriving an inspiration from the above illustrated entombment methodologies. Writings such as the Thevaram (7731) existing in the seventh century suitably refers to the graveyard as muthukadu, this has been found in another writing called as the Purananuru (356:4). Based on the above sayings, it is evident that the individuals are usually cleared out once they expire; on the other hand their internments alone remain as a lasting monument [2].

Various prevailing Indo-Aryan and Dravidian dialects such as Hindi, Bengali, Gujarati, Tamil and Kannada that are preferred as the conversation languages by the people in India today have received their contents from are spoken in present-day India have their contents gotten from Brahmi. Brahmi not only serves as the mother language offering contents to the various other languages as mentioned above, it also serves as a common language in the neighbouring countries like Sri Lanka, Myanmar, Thailand, and Tibet, among the others. Kanas has been viewed as an insignificant unit in the writing framework of Japan, even this unit has been believed to be derived from the mother language Brahmi. Similarly Aksara appears to be the insignificant unit in Brahmi and is a hypothetically intriguing segment of the syllable, that is, (consonant) (consonant) (consonant) (vowel) or just the last consonant. This unit is suitably decrypted as far as the current models of the syllable structure in the generative phonology are concerned. The outdated etymological idea of the Aksara unit has been found to yield money in the present scenario [3].

An accord possessed by the Brahmi impressions and Brahmi potsherds was to imprint Graffiti Marks either at the end or at the centre. A similar pattern is seen on the Tisamaragama Brahmi potsherd. In general the Brahmi engravings collected on the tunnels take a form from the left to a specific side possessing certain special cases wherein a couple of engravings in Tamil Nadu and Sri Lanka are found to compose the same from the privilege to one side. The existence of the privileges on the left side of the engravings might be because of the tall and unreachable upper edges; this reason would have led them to imprint the contents from the upper portions to the lower portions of the cavern [4].

The above mentioned limitations could have been eliminated by incorporating ceramic type materials for accomplishing the writing procedures as there are no motivations for maintaining the engravings from left to right and for collecting new engravings from the appropriate locations towards the left. The engravings on ceramics do not follow a specific pattern; the engravings in them are not imprinted on the left as there are no confirmations to demonstrate the double patterns of writing engravings on the stoneware. Preferably the three letters are scrutinized as "Tiṛaḷi" (jpusp) from the proper location to the left and can also be studied as "Puḻaiti" (Giojp) from the left to the right. The principal letter sounding as "Pu" possesses (G) a certain level of twist to one particular side on the straight line, preferably portraying the correct side. The above nature of the letter "Pu" does not resemble to the one present on the Brahmi type of cavern engravings. The above illustrated engraving types are present in the stoneware's found in the counties of Sri Lanka and Tamil Nadu, suitably bearing the Brahmi scripts.

From the obtained wide variety of Sangam hoardings the success of these engravings has been confirmed. The sonnets of the Sangam hoardings are as follows, the kuruntokai, ainkurunrru, Purananuru, patirruppattu and tolkappiyam in its language structure, preferably dated between the second B.C. [5]. These features prove that the Tamil language possesses a very refined form of writing framework during the above illustrated time span. Another visible confirmation is the characterization of the word eruttu in Tolkappiyam, this word comprises of thirty letters in a prescribed sequence from a to n together with the three other auxiliary structures.

எழுத்தெனப்படுப
eṛuttenappaṭupa
அகரமுதல் னகர இறுவாய் முப்பஃதென்ப
akaramutal nkara iṛuvāy muppaxtenpa
சார்ந்துவரன் மரபின் மூன்றலங் கடையே
cārntuvaran marapin mūṉṛalaṅ kaṭaiyē

Mahadevan (1966:58), has presented a letter set arrangement of Tamil-Brahmi based on a compilation of engravings, this revealed the usage of the Tamil-Brahmi contents in various periods (See Figure 1). It reveals the fact that the Tamil-Brahmi contents incorporated for use during the principal Thousand years B.C. neither made use of the dab, nor it recognized the short and the long e and o by the methodology reserved for the identification of a certain type of symbols.

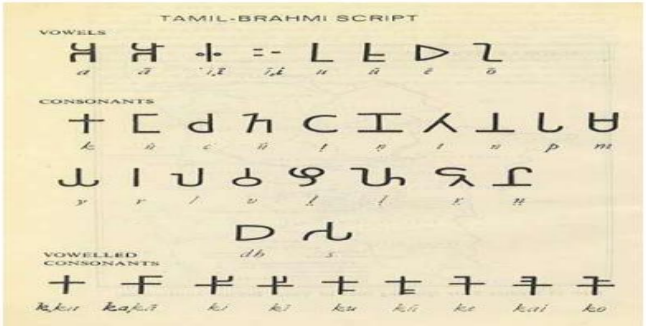

Figure 1: Tamil- Brahmi Alphabet reconstructed depending on different inscriptions found in Tamil Nadu (Mahadevan 1968: 56)

The most remarkable viewpoint in achieving a high level character acceptance for the ancient Tamil is to perfectly discriminate the suitable component extraction methodology. The above mention viewpoint appears to be a difficult achievement in the research field focussed on the various image processing techniques similar to that of the pattern acknowledgments. By means of achieving the above mentioned feature the automation procedure can be progressed together with the enhancement in the interface among man just as a machine in various applications. Various late investigations are focusing on new strategies furnished with facilities that can essentially enhance the accuracy level of the system with a minimized computational time. The common grouping of the character acknowledgments is thus isolated just as an on-line acceptance strategy. As far the previous acknowledgment strategy is concerned it has been observed that the method makes use of a digital camera for capturing the input image, however in the previous extractions two dimensional directions of back to back focuses have been illustrated as a component of time similar to that of the essayist's organization of strokes. The strategy of discriminating the existing letters on the ancient stone engravings as words is the idea behind the Offline ancient Tamil text acknowledgment technique. Further, the neural networks have been incorporated and utilized for the purpose of achieving an enhanced level of acknowledgment accuracy in the disconnected strategies.

Identifying new systems for upgrading the acknowledgement accuracy features of the ancient Tamil Character is the objective of the current research works. The pre-processing step has been considered as an observable advancement in the sequence of recognizing the ancient Tamil language. Once the pre-processing step ceases the segmentation procedure would get into action, this procedure is similar to the extraction process. The task of the pre-processing process is to perfectly shape the input image into an appropriate structure for performing the required segmentations. This segmentation procedure adopts a portion of the input image and shapes them into individual characters by means of resizing each of them into m x n pixels appropriately for the purpose of constructing a network. Different types and forms of element extraction strategies are available for discriminating the ancient Tamil characters. At the other end the task of grouping and recognizing the characters is accomplished by the artificial neural networks. The reason for adopting the neural networks in the above mentioned procedures is because of its enhanced acknowledgment precision for the offline ancient Tamil character acknowledgment framework, it is considered as the most solid apparatus with high speed [5]. Gradient Boosting tree is used for classifying the shape and Hough transforms incorporated for the feature extraction strategies. Section 2 presents the literature review. The introduced methodology and the obtained experimental results are therefore discussed in the sections 3 and 4 respectively. Finally the conclusion of the introduced work is thus presented.

II. RELATED WORK

Rajakumar and Bharathi [6] have introduced the contour-let change; this is a strategy that helps in the recognition of the Tamil characters from the stone engravings. Previous techniques associated with the Wavelet changes have insisted its inefficiency in reconstructing the bended images. This prescribed limitation can be very well adjusted by incorporating the contour-let change strategy. The wavelet change associates to a two dimensional type of methodology whereas the Contour-let change associates to a three dimensional type of methodology. The characters from the input images can be recognized by the Clustering system; on the other hand the fuzzy median filters can be utilized for bestowing the noise factors. Neural networks are mainly incorporated for the purpose of obtaining an accurate acknowledgment of the Ancient Tamil characters; this is because the neural networks possess the ability of appropriately preparing and contrasting the data and the current century characters.

Mahalakshmi and Sharavanan [7] familiarized a basic methodology for the purpose of recognizing and explaining the Tamil engravings. Apart from the above introduction a step by step study was accomplished on the ancient Tamil engravings together with the proposal of the current century characters. Lab VIEW was adopted and used for recognizing and decrypting the Tamil stone engravings. For the task of sectioning the images of the ancient contents the Segmentation procedure was incorporated. The various types of segmentation systems utilized are as follows, The Particle Swarm Optimization (PSO) system, the Discrete PSO (DPSO) system and the Fuzzy PSO (FPSO) system. Up gradation of the images is accomplished by the contour let change technique, the noise factors in the images are eliminated by the fuzzy median filters. Re-enactment of the suggested methodology was performed under Matlab and Lab VIEW.

Rajakumar and Bharathi [8] has suggested that the task of recognizing the ancient Tamil character is a noteworthy zone of research and have further discovered various types of applications in the pattern acknowledgment theory. It appears to be disapproving for accepting various types of techniques for automating the methodology of inputting the characters for the various types of incidents. The author has been found to introduce an ancient Tamil character acceptance agreement suitably obtained from an artificial resistant. This computation suggests resisting the organic standards and according to this the character acceptance rate has been enhanced with a diminished acceptance developing time. The introduced methodology was thus rebuilt and the obtained results confirmed that the methodology possessed an enhanced speed level when compared with the conventional techniques meant for acknowledging the ancient Tamil characters obtained from the neural networks. The agreement has been found to possess an adaptable learning feature similar to that of a resistant memory system such as that of a biologically resistant structure. This is connected for the purpose of discriminating the abnormalities similar to the ones in the pattern recognitions.

Rajakumar and Bharathi [9] have introduced a creative component removal methodology possessing features for enhancing the results obtained from the comparison of two different casted ancient Tamil characters that have been selected for acceptance. In this work, the author has made use of occasions of Tamil characters from 6 unmistakable hundreds of years.

The technique has its premise in F-ratio, a measurement decided by the ratio between-class just as within the class change. F-ratio modifies the attribute vector of 2 nearly similar to that of the shape of the characters by means of measuring the features portions. These measuring strategies have assisted in the determination of characters possessing look alike comparative features. This proposed strategy has been found to work well by effectively enhancing the element components having a place with the proved pieces of the developed characters by means of limiting the various feature segments of the run of the mill bits of characters. The introduced methodology has been observed to make use of the angle attributes and the template coordinating methodologies, here an acceptance precision rate of around 94% is thus obtained.

Rajakumar and Bharathi [10] have performed various forms of research on the structural attribute that are useful in the offline acceptance of the Ancient Tamil characters. It is quite ridiculous to expect the grouping procedures of the characters to happen with the help of the Structural attributes alone. Hence there appears a need for the existence of a certain type of features together with the artificial neural networks for the purpose of enhancing the presentation of the concerned structure. This introduced agreement essentially with holds the necessary features for obtaining a precision rate of around 97.9% for certain type of letters at a normal rate of around 80% and furthermore as far as the time utilization factors are concerned.

Kumar and Poornima [11] have proposed the epigraphic engravings. This presumes a huge job in identifying the educated past and in aggregating the characters possessing a particular place in various time periods. This introduced structure possesses the ability of scrutinizing the ancient Tamil characters possessing a place in various time periods by means of testing a limited quantity of characters named as the inspected characters in Tamil. Through the computerized implicit, the verified characters would be essentially isolated from the content just as the ones composed with the characters that occupy a prescribed location in distinctive periods using the machine information. From the above illustrations it is clear that the suggested framework is composed of various modules such as the image acquisition module, the binarization module, pre-processing attributes and the features extraction module. The segmentation process is thus observed to be similar to that of the expectation periods, thus accomplished by making use of the Transductive Support Vector Machine (TSVM). The re-enactment output illustrates the attainment of an enhanced level of precision in comparison with the precision levels of the Support Vector Machine (SVM).

Tomar et al [12] introduced a mixture of strategies to naturally discriminate the various characters and recognition of the concerned contents. This configuration was thus observed as a dense packet of contents made up of techniques related to the image processing methodologies, segregation techniques for the discrimination of the system attributes and was further made up of certain removal strategies and methodologies for minimizing the dimension based attributes. In spite of a wide research in the area of ancient engravings the task of character acceptance has been observed as a tedious process, hence there appears a need for discovering successful procedures for the accomplishment of the same. This survey can be observed as the reason for establishing preliminary dimensions in the area of image pre processing and hence the dimensionality minimization strategies have been found as the highlight and arrangement of the entire procedure.

III. PROPOSED METHODOLOGY

Figure 2 portrays the introduced methodology's architecture. In this method the noise elimination task is essentially encountered by the median filter. Presence of noise factors in an image can be essentially viewed as grains and it is these grains that lead to unpredictable modifications in the strength of an image. Noise factors are usually observed during the image acquisition process. Noise factors can be eliminated by incorporating strong algorithms capable of removing them completely else suitable filters can be adopted for the same. Different types of noise factors are viewed in the document images, few of them are as follows, Salt and Pepper noise, Gaussian noise, Gamma noise, Uniform noise etc. Apart from the above mentioned strategies various kinds of filtering mechanisms can also be incorporated for the removal of the noise factors, these are as follows, the Gaussian filtering method, Min-max filtering method etc. One of the filtering methods adopted for eliminating the salt and pepper noise is the Median filter.

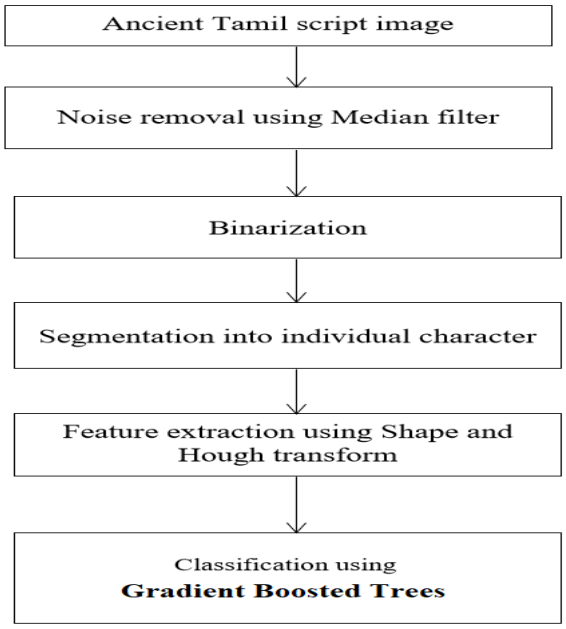

Figure 2: Proposed work flowchart

One of the most common filtering method adopted in the digital filtering methodology is the Median filtering technique, this technique has been found to safeguard the image edges during the noise elimination process under certain defined conditions.

It is a nonlinear type of local filter in which the output magnitude essential illustrates the middle component of a sequenced array of pixel values from the filter windows. The working procedure of the Median filter is that it perfectly filters all the pixels contained in the image in turns, followed by which the determination of the surrounding representative would be accomplished by the closest neighbour.
Another function adopted by the Median filter is that it would substitute the magnitude of the pixel with the median of the selected values.

The initial step here is to arrange the neighbouring values in a numerical order, at the same time the value of the pixel under observation would be deputed with a middle (median) pixel value.

The neighbourhood in this procedure is named as the window. It is around the target pixel that the various shaped window pieces would be centred, as far as the window defined for a two dimensional image is concerned a perfect square shape is selected. The middle value possessed by the neighbouring pixels is mostly the value of a pixel in the neighbourhood within the window. Hence the reason for the creation of a novel impractical value by the median filter, this creation usually occurs infrequently. The median filtering method is thus observed as a superior strategy than that of the mean filtering method, this is because of the safeguarding procedure adopted by the method in terms of the sharp edges. One of the major limitations of the median filtering methodology is the presence of an artificial output which appears as a median value in the window. Most of the available filtering methodologies are found to discard the noise factors present on smooth patches or on the areas of a signal, the limitation is that it fails to eliminate the noise sources present on the edges, as a result of which the image gets affected during the process. As far as the text removal and identification procedures are concerned it turns out to be mandatory for eliminating the noise sources from the image, further it insists the need for safeguarding the edges as well. Protection of the edges is essential as they are the ones involved in the visual appearance of the images. The above mentioned aspects turn out the median filtering procedure as the most preferred methodology in digital image processing.

Segmentation: It is in this process that the separation of the individual characters from the prehistoric text happens. This is accomplished by incorporating the projection profile analysis strategy and the component labelling strategy in coordination. The stages that make up the segmentation process are as follows:

- Line Segmentation
- Word Segmentation
- Character Segmentation

Segmentation process fragments a single text line from the scanned documents, single word from a single line as well as a single character from the single word. Two major classes make up the segmentation process, they are stated below:

a) External segmentation is performed for separating the paragraphs, single lines or words.

b) Internal segmentation is performed for separating the individual characters.

Various methods are incorporated for fragmenting and isolating the individual characters suitably obtained from the projection profiles, connected component labelling or from the white space and pitches respectively.

Features Extraction: The performance of a character recognition system is thus determined with respect to a certain type of attributes contained in the process. The likelihood of attaining an enhanced level of recognition depends on the selection of the most suitable feature selection methodology. Unique determination of the concerned character sets turn out to be feasible with the extracted features and also there appears a need for the existence of a huge number of variations among the features that belong to a specific character set.

Shape Transform Recognition: The basic idea of the strategy is to consider the position of an individual pixel as an attribute and that it must not be eliminated at any point. The Shape Transformation protocol holds features similar to that of the dynamic programming strategy, adopted for the accomplishment of the lexicographical correction task which is observed as a post processing stage in OCR. The binary character image x may be converted into image y through:

Leaving particular pixels which are unspoiled

Moving some x pixels into locations that are relative to its respective closer pixel within and

removing some pixels when needed.

Because of variances in field, deletion, substitution or Insertion matrices is not present. Hence, it is important to find our own. Towards substitutions, there exist a absence of cost attribute as it is similar in holding it in circumstances as they are general. The included costs in pixel shifting to a various location is recognized equivalent to Euclidean distance between destination and source locations. in x, as each pixel is either translated or replaced, the corresponding x pixels and y are relative. The relative ones are not reusable.

Hough transform: In image processing, for extracting feature are employed. The statistical technique comprise main significance in present time in classification domain and it is highly important for OCR. It gives a superior outcomes, when we merge Hough transform and statistical technique for facial and font recognition. In computer vision field, the Hough Transform was employed actually to recognize lines in images. It was employed for different geometrical attributes such as ellipse and circle. It was improved in eighties in order that it is appropriate for standard shape detection which gains GHT. Through the global purview, the HT cause for successful is provided. There is no intention to access the fundamental knowledge over point distributions but, the voting process of entire points tend to increase in accumulators. The robustness is ensured by voting process to HT for missing edge point. The entire points which is derived individually is useless but each point polls for a certain shape.

Image points (x, y) sequence which rely on a straight line might be given by relationship, f, as

$$f\big((\hat{m},\hat{c}),(x,y)\big) = y - \hat{m}x - \hat{c} = 0 \qquad (1)$$

Variables are represented through m and c, intercept and slope which categorized the line in Eq (1) map entire variable combinations (mˆ,cˆ)rates towards image points sequence.

3.1 Classification with gradient tree boosting

A gradient tree boosting is trained in end phase by employing XGBoost with the entire derived features from prior phase.

GBT with many other techniques of tree ensemble learning had been employed extensively in competitions of data mining. It is highly variant to input scaling and it might learn from high order interaction among features. The GBT is trained by additive way unless from other techniques of tree ensemble. It construct other tree at every time t, to reduce present residual model.

The main function might be discussed as

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y_i}^{t-1} + f_t(x_i)\right) + \Omega(f_t), (2 \qquad (2)$$

Let loss function is denoted through l which measures the variances among the i-th instance label and prediction at final phase in addition to present tree output; and regularization term is denoted through Ω(f_t) which the complexity of novel tree is penalized. One significant and well-known execution in GBT is XGBoost because of it huge efficacy and success in different competitions. Through default, it manages the missing values. From training data, optimal default direction learned by the method in every tree node. The instance would be segmented in default feature direction when the feature value is missing.

As training data only comprise of 325 gold labels notes and are imbalanced over various severity scores highly, we executed with resampling notes to enhance the data volume in training and manage the note counts with various scores. We constraint further the higher decision trees depth and minimal child weights, and we provide randomness towards training procedure through sampling training and features samples in every phase to prevent overfitting. To adjust the entire classifiers hyper parameters, we have employed 5-fold cross-validation. With the similar environmental setup, we have trained additionally the techniques like decision tree (DT), random forest (RF), logistic regression (LR), by employing scikit-learn library over the GBT classifier.

IV. EXPERIMENTAL EVALUATION

The experimental examination is carried with the feature selection and classification. We employ 9 ancients with every letter comprising 35 instances every experiment.

Table 1 Performance analysis of various classifier in terms of Accuracy

Classifier Accuracy	J48	KNN	NN	GBT
Shape Features	88.57	86.03	90.48	91.49
Hough Features	89.84	87.94	91.75	92.80
Concatenation of features	91.11	88.89	93.02	94.05

It might be examined from fig. 7 which GBT classifier enhanced accuracy when compared to other techniques. For concatenated features, GBT classifier enhanced accuracy for concatenated features through attaining 94.05% when compared to KNN which attains 88.89 and J48 and NN classifier attains 91.11% and 93.02%. Hence, by means of accuracy, the proposed GBT classifier is superior.

.

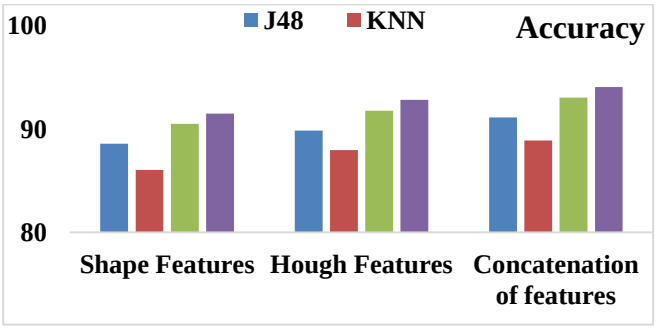

Figure. 4 Comparison outcomes in terms of accuracy

Table 2 Performance analysis of various classifier in terms of Precision

Classifier Accuracy	J48	KNN	NN	GBT
Shape Features	0.87	0.87	0.91	0.91
Hough Features	0.89	0.88	0.91	0.91
Concatenation of features	0.90	0.87	0.92	0.94

Fig. 8 shows the proposed GBT classifier efficiency in terms of precision when compared to other techniques. For concatenated features, GBT classifier attains enhanced precision for concatenated features through obtaining 0.94% when compared to KNN which attains lowest precision rate of 0.88% and J48 and NN classifier attains 0.91% and 0.93%. Hence, by means of precision, the proposed GBT classifier is efficient.

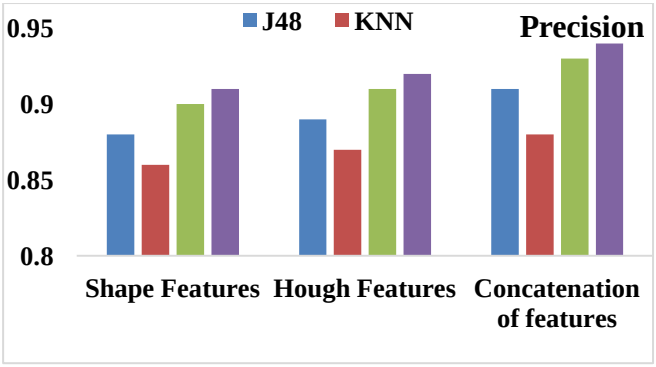

Figure. 5 Comparison outcomes in terms of precision

Table 3 Performance analysis of various classifier in terms of Recall

Classifier Accuracy	J48	KNN	NN	GBT
Shape Features	0.87	0.87	0.91	0.91
Hough Features	0.89	0.88	0.91	0.91
Concatenation of features	0.90	0.87	0.92	0.94

Fig. 8 demonstrates the proposed GBT classifier efficiency in terms of recall when compared to other techniques. For concatenated features, GBT classifier obtains improved recall rate of concatenated features through obtaining 0.94% when compared to KNN which attains lowest precision rate of 0.88% and J48 and NN classifier attains 0.91% and 0.93%. Therefore, by means of recall, the proposed GBT classifier shows effective results.

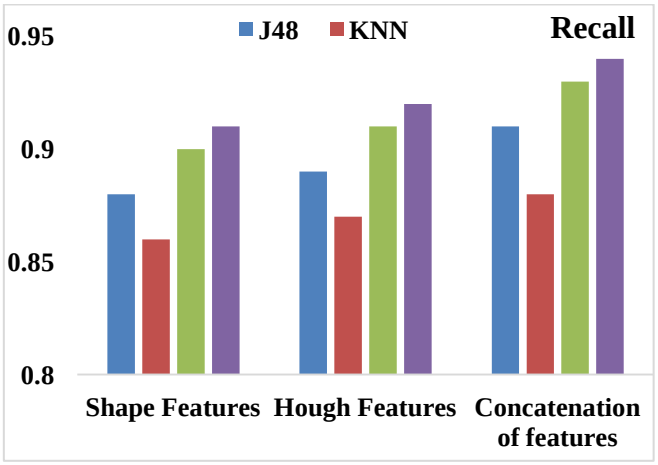

Figure. 6 Comparison outcomes in terms of recall

It is examined from the above discussion that the neural network, J48, KNN classifier shows reduced performances in classification of ancient tamil script. And it is obsolute that the projected GBT classifier shows enhanced results.

V. CONCLUSION

For Tamil script optical character recognition, this paper provides a classification and feature extraction techniques analysis. For removing the noise, median filter is employed and it segment every distinct character for extraction of feature by employing shape and Hough transformation. Three classifiers like neural network, J48, KNN are compared in this work. The outcomes are compared and enhancements are noted in result validation. For concatenation of features, enhanced recognition accuracy 94.05% is reached through the projected method. It is clear from the outcomes that the GBT classifier is enhanced when compared to other methods. For recognition accuracy maximization, the work might be extended through inserting additional relative features.

REFERENCES

1. Preethi P. and Mamatha H.R., A Review on Automation of Ancient Epigraphical Images, International Journal of Database Theory and Application, 9(4), 143-150 (2016)
2. Iniyan E., Burial and Fuunerary Culture of Ancient Tamils (During 1000 BC-250/300 AD), International Journal of Social Science and Humanity, 5(12), 1068 (2015).
3. Patel P.G., Constructing a Framework for theories of the Brahmi Writing System.
4. Pushparatnam P., Tamil Brahmi Inscription Belonging to 2200 years ago, Discovered by German Archaeological Team in Southern Sri Lanka, Jaffna University International Research Conference (2014).
5. Rajakumar S. and Bharathi V.S., 12th Century Ancient Tamil Character Recognition from Temple Wall Inscriptions. iManager's Journal on Embedded Systems, 1(2), 27 (2012).
6. Rajakumar S. and Bharathi V.S., Century Identification and Recognition of Ancient Tamil Character Recognition, International Journal of Computer Applications, doi: 10.5120/3090-4237 (2011).
7. Mahalakshmi M. and Sharavanan M., Ancient Tamil script and recognition and translation using LabVIEW. In Communications and Signal Processing (ICCSP), 2013 International Conference, IEEE, 1021-1026 (2013).
8. Rajakumar S. and Bharathi V.S., Eighth century Tamil consonants recognition from stone inscriptions, In Recent Trends In Information Technology (ICRTIT), 2012 International Conference, IEEE, 40-43 (2012).
9. RajaKumar S. and Bharathi V.S., Similar Shaped Ancient Tamil Character Recognition using Gradient Feature and Template Matching, Digital Image Processing, 4(6), 337-340 (2012).
10. Rajakumar S. and Bharathi V.S., Offline Ancient Tamil Character Recognition System Based On Structural Features, iManager's Journal on Communication Engineering and Systems, 1(3), 17 (2012).
11. Kumar S.V.K. and Poornima T.V., An Efficient Period Prediction System for Tamil Epigraphical Scripts Using Transductive Support Vector Machine, International Journal of Advanced Research in Computer and Communication Engineering, 3(9), 7999-8002 (2014).
12. Tomar A., Choudhary M. and Yerpude A., Ancient Indian Scripts Image Pre-Processing and Dimensionality Reduction for Feature Extraction and Classification: A Survey, International Journal of Computer Trends and Technology (IJCTT), 21(2) (2015).

AUTHORS PROFILE

T.S.Suganya is a Research Scholar in the Department of Computer Science and Engineering in Sathayabama University, Chennai. She received her M.C.A degree in computer Applications and M.Phil. in computer science. Her interest includes Image Processing and Pattern Recognition.

Murugavalli Subramaniam is presently working as professor and Head in Panimalar Engineering College, Chennai. She received her Bachelor degree (1995) in Electronics and Communication Engineering ,Thiagarajar College of Engineering, Madurai Kamarajr University, Madurai, Master's (2000) in Computer Science and Engineering ,Government College of Engineering, Manonmaniam Sundaranar University, Thirunelveli and Doctorate (2009) in Anna University, Chennai. Her area of interest is Simulation of Image Processing, Soft Computing, Medical Imaging, Genetic algorithm Fuzzy Logic, Mobile and Grid Computing. She is a life member of CSI and ISTE and Annual member of IEEE. She published papers in international (35) and National (2) journals, presented papers in National and International conferences (40) and published three books. She received Best international paper presenter Award and Maximum Publication in CSIC Award from Computer Society of India (CSI).