HIGHLIGHTED ARTICLE

DOI: 10.34256/irjt21210

INTERNATIONAL RESEARCH JOURNAL OF TAMIL
சர்வதேசத் தமிழ் ஆய்விதழ்

OPEN ACCESS

பனை ஓலைச்சுவடிகளில் எழுதப்பட்டுள்ள தமிழ் எழுத்துக்களை கணினி வழி அடையாளப்படுத்துதல்

இரா. ஸ்பர்ஜன் ரத்தீஷ் [அ, *], மு.முகம்மது சாதிக் [அ]

[அ] சதக்கத்துல்லாஹ் அப்பா கல்லூரி, (மனோன்மணியம் சுந்தரனார் பல்கலைக்கழகத்துடன் இணைவு பெற்றது), திருநெல்வேலி-627011, தமிழ்நாடு, இந்தியா.

Tamil Character Recognition in Palm Leaf Manuscripts

R. Spurgen Ratheash [a,*], M. Mohamed Sathik [a]

[a] Sadakathullah Appa College, (Affiliated to Manonmaniam Sundaranar University), Tirunelveli-627011, TamilNadu, India.

* Corresponding Author:
spurgen@gmail.com

Received : 31-01-2021
Revised : 15-02-2021
Accepted : 21-02-2021
Published : 26-04-2021

Check for updates

ABSTRACT

Tamil characters are of historical significance. The shapes of the character and its writing continued to be changed by several reformers in each century until the period of the nineteenth century. The Tamil characters in the palm leaf manuscripts are based on the author's writing style and strokes in use at the time. Reading palm leaf manuscripts is a challenge for the modern generation who are unaware of the character strokes and writing patterns written in earlier times, and the younger generation neglects to read and understand the contents written in palm leaf manuscripts. To read Tamil palm leaf manuscripts it is necessary to remember the shapes of the character that has changed over time and to compare and recognize the characters. This research paper explains how to recognize the Tamil characters written in palm leaf manuscripts by computer. By this research, the Tamil characters can be compared with different strokes and shapes and the exact character can be recognized accurately and quickly.

Keywords: Tamil manuscripts, Tamil character, Text line segmentation, Character segmentation, Character recognition

ஆசிரியர் குறிப்பு

இரா. ஸ்பர்ஜன் ரத்தீஷ், சதக்கத்துல்லாஹ் அப்பா கல்லூரியின் ஆராய்ச்சி மாணவர் ஆவார். இது இந்தியாவின் தமிழ்நாட்டில் உள்ள திருநெல்வேலி மனோன்மணியம் சுந்தரனார் பல்கலைக்கழகத்துடன் இணைக்கப்பட்டுள்ளது. எண்மின் மயமாக்கல் பட செயலாக்கம், ஆவண பட பகுப்பாய்வு மற்றும் தமிழ் மொழி பனை ஓலை கையெழுத்துப் பிரதிகளின் எழுத்து அங்கீகாரம் ஆகியவை அவரது முக்கிய ஆராய்ச்சி ஆர்வங்களில் அடங்கும்.

முனைவர் மு. முகமது சாதிக், இந்தியாவின் திருநெல்வேலியில் உள்ள சதக்கதுல்லாஹ் அப்பா கல்லூரியின் முதல்வராக உள்ளார். இந்தியாவின் திருநெல்வேலியில் உள்ள மனோன்மனியம் சுந்தரனார் பல்கலைக்கழகத்தில் கணினி அறிவியல் மற்றும் கணினி அறிவியல் மற்றும் தகவல் தொழில்நுட்பத்தில் தேர்ச்சி பெற்ற இரண்டு முனைவர் பட்டங்களைப் பெற்றார். அவர் பல தேசிய மற்றும் சர்வதேச கருத்தரங்குகள், மாநாடுகளில் கலந்து கொண்டு ஏராளமான ஆய்வுக் கட்டுரைகளை வழங்கியுள்ளார். பல சர்வதேச பத்திரிகைகளில் வெளியீடுகளுடன், 40 க்கும் மேற்பட்ட ஆராய்ச்சி

அறிஞர்களுக்கு வழிகாட்டியதைத் தவிர இரண்டு புத்தகங்களை வெளியிட்டுள்ளார். இந்தியாவின் தமிழ்நாட்டில் உள்ள பல்வேறு பல்கலைக்கழகங்கள் மற்றும் தன்னாட்சி கல்லூரிகளின் பாடத்திட்ட மேம்பாட்டுக் குழுவில் உறுப்பினராக உள்ளார். மெய்நிகர் உண்மை, எண்மின் மயமாக்கல் பட செயலாக்கம் மற்றும் உணரி பிணையம் ஆகியவை அவரது ஆராய்ச்சியின் சிறப்புப் பகுதிகள்.

முன்னுரை

உலகின் செம்மொழிகளில் தமிழ் மொழியும் ஒன்றாகும். தொன்மை வாய்ந்த இம்மொழி ஆசிய கண்டத்தில் குறிப்பாக இந்தியா, சிங்கப்பூர், இலங்கை ஆகிய நாடுகளில் ஆட்சி மொழியாகவும். மேலும், மொரீசியஸ் போன்ற பிற கண்டங்களில் உள்ள நாடுகளில் கலாச்சார மொழியாகவும் ஏற்றுக்கொள்ளப்பட்டுள்ளது. முதன்முதலில் முன்னோர்களால் ஒலி வடிவில் தொடங்கி பின் குறியீடுகளால் எழுதப்பட்டு அதன்பின் ஓலைச் சுவடிகளிலும், கற்களிலும் பதிக்கப்பட்ட தமிழ் எழுத்துக்கள் தற்போது உருவத்தில் அநேக மாற்றம் பெற்று உயிரெழுத்துக்கள் பன்னிரெண்டும், மெய்யெழுத்துக்கள் பதினெட்டும் ஆயுத எழுத்து ஒன்றும் சேர்ந்து மொத்தம் 247 ஆக எழுத்து, எண், பேச்சு மற்றும் எழுத்து வழக்கிலும், கணிப்பொறிப்பதிப்பிலும் பயன்படுத்தப்பட்டு வருகிறது. பழங்காலத்தில் எழுதப்பட்ட பனை ஓலைச் சுவடிகள் அறிவியல், மருத்துவம், புவியியல், சோதிடம், இலக்கியம், நாடகம் என அனைத்து வகையிலும் முன்னோர்களின் அறிவாற்றலால் எழுதப்பட்ட அறிவுப் பொக்கிசமாக திகழ்கிறது. பனை ஓலைச்சுவடிகள் என்பது பனை மரத்தின் ஓலைகளை நறுக்கி பல நிலைகளில் பக்குவப்படுத்தி அதில் 5 முதல் 6 எழுத்து வரிகளை இரும்பினால் செய்யப்பட்ட ஊசியால் எழுதி 40 முதல் 50 ஓலைகளை ஒன்றாக கட்டி வைத்துள்ள தொகுப்பு ஆகும். அவை படம் 1-இல் காட்டப்பட்டுள்ளன. இந்த அறிவுப் பொக்கிசத்தை பாதுகாப்பது மிகவும் கடினமானதாகவும், அதிக செலவினங்களை ஏற்படுத்தக்கூடியதாகவும் மற்றும் இடத்தை அதிகம் ஆக்கிரமிப்பதாகவும் உள்ளது. எண்மின்மயமாக்கல் (Digitization) என்ற தொழில் நுட்ப முறைப்படி அநேக ஓலைச்சுவடிகள் வருடிநகலி (Scanner) உதவியுடன் படமாக்கப்பட்டு கணினியில் சேமித்து வைப்பதன் வழியாக எளிதாக கையாளவும், அனுப்பவும், பலமுறை பார்த்தாலும் பழுதாகமலும், பார்ப்பதற்கு ஏற்ப பெரிதாக்கவும், பராமரிப்பு செலவில்லாததாகவும், குறைந்த அளவு நினைவக இடத்தில் அதிக அளவு படங்களை சேமிக்கவும் முடிகிறது. ஆனாலும் அதில் எழுதப்பட்டுள்ள ஒரு சில எழுத்துக்களின் தோற்றம், எழுத்து முறை தற்போது உள்ள தமிழ் எழுத்துக்களை விட வேறுபட்டிருப்பதாலும், எழுத்துக்கள் ஒன்றையொன்று ஒட்டிக் கொண்டும், வார்த்தைகளுக்கு இடையே இடைவெளி இல்லாததாலும், அதனை வாசித்து அர்த்தம் விளங்கிக் கொள்ளவதில் சிரமம் ஏற்படுவதாலும் எண்மின்மயமாக்கல் பயனற்றதாய் போகும் அபாய நிலைக்குத் தள்ளப்பட்டுள்ளது. ஓலைச்சுவடிகளின் எழுத்துக்களை வாசிப்பது அதற்கென பிரத்யேக பயிற்சி பெற்றவர்களால் மட்டுமே சாத்தியமாகிறது.

படம் 1 ஓலைச்சுவடிகள்

இதனால் அநேக ஓலைச்சுவடிகள் புத்தகங்களாகப் பதிப்பிக்கப்படாமல் அழியும் நிலை நிலவி வருகிறது. இந்த ஆராய்ச்சி பனை ஓலைச் சுவடிகளில் உள்ள தமிழ் எழுத்துக்களை கணினியை கொண்டு அடையாளப்படுத்தி அதனை கணினி தட்டச்சு செய்யப்பட்ட எழுத்துக்களாக காட்சிப்படுத்துவதைப் பற்றி விளக்குகிறது. இதனால் ஓலைச்சுவடிகளில் உள்ள தமிழ் எழுத்துக்களை

வேகமாக வாசிக்கவும், மிகச் சுலபமாக புரிந்து கொள்ளவும் முடியும் என்பதை இந்த ஆராய்ச்சி கட்டுரை விளக்குகிறது.

உரையாசிரியர்களின் விளக்கங்கள்

தமிழ் எழுத்துக்கள் எழுதப்பட்ட முறையும், பனை ஓலைச்சுவடிகளில் எழுதப்பட்டுள்ள தமிழ் எழுத்துக்கள் கொண்டுள்ள பல்வேறு வகையான சவால்களும், ஓலைச்சுவடிகளில் பயன்படுத்தப்பட்டுள்ள எண்களின் முறைமையையும் விளக்குகிறது (Kattalai Kailasam, 2019). கையால் எழுதப்பட்ட தமிழ் எழுத்துக்களை கணினி வழி அடையாளப்படுத்துவதில் சுருளல் நரம்பணுப் பிணையம் (Convolutional Neural Networks) முறை பிரயோகப்படுத்தப்படுவது பற்றி குறிப்பிடுகிறது (Deepa et al., 2019). ஒரு நபரால் எழுதப்பட்ட பனை ஓலைச் சுவடிகளில் உள்ள தமிழ் எழுத்துக்களை சுருளல் நரம்பணுப்; பிணையம் வழியாக அடையாளப்படுத்தப்பட்டு உள்ளது (Sabeenian et al., 2019). பூலியன் அணிக்கோவை (Boolean Matrix) மற்றும் மூச்சி முதல் தேடல் (Breadth First Search) நுட்பத்தைப் பயன்படுத்தி ஓலைச்சுவடிகளில் உள்ள தமிழ் எழுத்துக்களை கணினி வழி அடையாளப்படுத்தி உள்ளனர். அதில் சாய்ந்த எழுத்துக்கள் போன்ற அடிப்படை பிரச்சனைகளை கையாளுவதில் உள்ள சிக்கல்கள் விளக்கப்பட்டு உள்ளது (Vellingiriraj, & Balasubramanie, 2014). கையால் எழுதப்பட்ட தமிழ் எழுத்துக்களை அடையாளப்படுத்துவதில் செயற்கை நரம்பணுப்பிணையம் (Artificial Neural Networks) முறை பயன்படுத்தப்பட்டு உள்ளது (Banumathi, & Nasira, 2011). ஆழ்கற்றல் (Deep Learning) மற்றும் பெரிய தரவு (Big Data) முறையில் கற்கும் திறன் புகுத்தி தமிழ் எழுத்துக்களை அடையாளப்படுத்துவதில் கருத்தாய்வு செய்யப்பட்டு உள்ளது (Kannan, & Subramanian, 2015). பிம்பவியல் (Image processing) வழியாக பிம்பம் மேம்படுத்துதல் (Image Enhancement), பிம்பம் பிரித்தல் (Image Segmentation) மற்றும் பிம்பம் மறுசீரமைப்பு (Image Restoration) ஆகிய நுட்பத்தைப் பயன்படுத்தி தமிழ் எழுத்துக்களை கணினி அடையாளப்படுத்துவதில் கருத்தாய்வு செய்யப்பட்டு உள்ளது (Challa, & Mehta, 2017). தென்கிழக்கு ஆசிய நாடுகளில் உள்ள பனை ஓலைச்சுவடி எழுத்துக்களை ஆவண பட பகுப்பாய்வு (Document Image Analysis)நுட்பத்தின் வழியாக அடையாளப்படுத்துவது பற்றி விவாதிக்கப்பட்டு உள்ளது (Kesiman et al., 2018). முதன்மை உபகரண பகுப்பாய்வு (Principal Component Analysis) நுட்பத்தை ஆழ்சுருளல் நரம்பணுப் பிணையம் (Deep Convolutional Neural Networks) நுடபத்துடன் இணைத்து கையால் எழுதப்பட்ட தமிழ் எழுத்துக்களை கணினி வழி அடையாளப்படுத்தப்பட்டு உள்ளது (Sornam, & Vishnu Priya, 2018).

முன்மொழியப்பட்ட ஆராய்ச்சி முறை

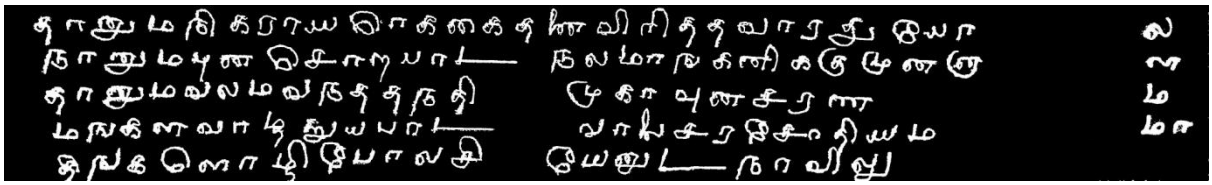
பனை ஓலைச் சுவடிகளில் தமிழ் எழுத்துக்கள் இரும்பாலான ஊசியைக் கொண்டு எழுதப்படுவதாலும், அகலம் குறைந்த ஓலைகளை மடியில் வைத்து எழுதுவதாலும் எழுத்து வரிசைகள் தட்டச்சு செய்தது போன்று நேர்கோட்டில் அமைந்து இருப்பதில்லை மற்றும் எழுத்துக்கள் ஒன்றை ஒன்று ஒட்டிக் கொண்டும், ஒன்றின் மேல் ஒன்று பிண்ணியும் காணப்படுகிறது. ஊசி கொண்டு எழுதப்பட்ட எழுத்துக்கள் பார்ப்பதற்கு தெளிவாக தெரிவதில்லை அதனால் அதன் மீது விளக்கு புகை, கரி ஆகியவை பூசப்படுகிறது. அவை படம் 2-இல் காட்டப்பட்டுள்ளது. இதனால் சரியாக பூசப்படாத இடங்களில் எழுத்துக்களின் தொடர்ச்சி விடுபடுவதால் கணினியில் அடையாளப்படுத்தும் போது எழுத்துக்களை உடைந்தாகக் கருதுகிறது. இதனை சரி செய்வதற்காக இந்த ஆராய்ச்சியில் பிம்பம் மேம்படுத்துதல் (Image Enhancement) நுட்பம் உபயோகப்படுத்தப்பட்டு உள்ளது.

படம் 2. அ. எழுத்தாணியால் எழுதப்பட்ட தமிழ் எழுத்துக்கள், **ஆ.** கரிபூசப்பட்ட எழுத்துக்கள்

சரியான எழுத்துக்களை ஒப்பிட்டு கணினி வழி அடையாளப்படுத்த ஒட்டிய தமிழ் எழுத்துக்களை பிரித்து தனி எழுத்து பிம்பங்களாக மாற்றுவது இன்றியமையாதது. இதற்காக எழுத்து வரிசை பிரித்தல் (Text Line Segmentation), எழுத்து பிரித்தல் (Character Segmentation), எழுத்து அடையாளப்படுத்துதல் (Character Recognition) என்ற படி நிலைகள் உபயோகப்படுத்தப்பட்டுள்ளது.

பிம்பம் மேம்படுத்துதல்

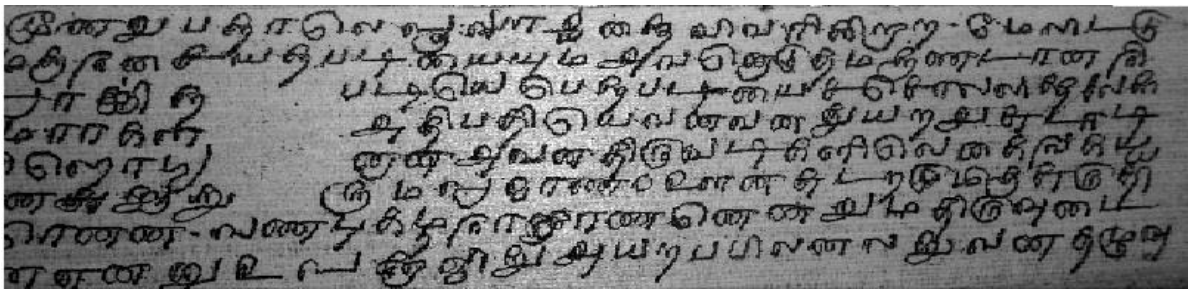
பனை ஓலைச்சுவடிகளின் தமிழ் எழுத்துக்களை அடையாளப்படுத்த சில முன் செயலாக்க முறைகள் பின்பற்றப்படுகின்றன. ஓலைச்சுவடிகள் வருடி நகலி மூலமாகவோ அல்லது எண்ணியல் நிழற்படக் கருவி (Digital Camera) மூலமாகவோ எண்மின்மயமாக்கப்படுகிறது. ஓலைச்சுவடிகளில் ஓலையின் நிறம் மற்றும் நாள்படிந்த கறைகளும் இருப்பதால் பின்னணி நீக்கத்தின் (background removal) போது எழுத்துக்களை தெளிவாக தெரியவிடுவதில்லை. இதனால் சுவடிகளின் மின்பிம்பங்கள் பிம்பம் மேம்படுத்தலுக்கு உட்படுத்தப்படுகிறது. அதன்படி எழுத்துக்கள் தெளிவைப் பெறுகிறது. அதன்பின் பின்னணி நீக்க முறையின்படி எழுத்துக்கள் வெண்மை நிறமாகவும், பின்னணியம் கறுப்பு நிறமாகவும் மாற்றப்படுகிறது அவை படம் 3-இல் காட்டப்பட்டுள்ளது. கடைசியாக உருவவியல் (Morphology) நுட்பத்தைப் பயன்படுத்தி எழுத்துக்களின் ஓரங்கள் மற்றும் அதன் கோடுகள் வலுவேற்றப்படுகின்றன. இத்தகைய நுட்பத்தினால் வலுப்பெற்ற எழுத்துக்களுடன் பிம்பம் அடுத்த கட்ட செயலாக்கத்திற்கு தயாராகிறது.



படம் 3 கருப்பு, வெள்ளை என இரு நிறங்களாக மாற்றப்பட்ட தமிழ் ஓலைச்சுவடி

எழுத்து வரிசையைப் பிரித்தல்

பனை ஓலைச்சுவடிகளில் எழுத்துக்கள் தட்டச்சு செய்வது போன்று சரியான வரிசையில் இல்லாமல் மேலும் கீழுமாக வளைந்து காணப்படும். ஒருசில நெடில் எழுத்துக்கள் நீண்டு கீழ் உள்ள எழுத்து வரிசையுடன் ஒட்டி காணப்படுகிறது. அவை படம் 4-இல் காட்டப்பட்டுள்ளது. தமிழ் எழுத்துக்களை கணினி வழி அடையாளப்படுத்த ஒவ்வொரு எழுத்தும் தனி படமாக இருப்பது அவசியமாகிறது. அதன்படி மேல் உள்ள எழுத்து வரிசையின் எழுத்துக்கள் கீழ் வரிசையுடன் ஒட்டியிருந்தால் அதனை பிரிப்பது இன்றியமையாததாகிறது.



படம் 4 எழுத்து வரிசை ஒன்றுடன் ஒன்று ஒட்டியிருத்தல்

பிரிக்காமல் இருக்கும் பட்சத்தில் அசலான எழுத்துக்கள் வேறொரு எழுத்துக்கள் போன்று தோன்றிவிடும். இதற்காக எழுத்து வரிசை பிரிப்பதற்கென இக்கட்டுரையின் ஆராய்ச்சியாளர்கள் பிரத்தியேகமாக எழுத்து வரி வெட்டுதல் வழிமுறையை (Text Line Slicing Algorithm) கண்டுபிடித்துள்ளனர். அதன்படி எழுத்துக்கள் நீண்டு வருமாயின் அது பிற எழுத்துக்களுடன் சேருமிடத்தில் பிரித்து விடுகிறது. இதனால் எழுத்துக்களின் வடிவங்கள் மாறுபடாமல் அசலான எழுத்துக்களுடன் எழுத்து வரிசை பிரித்தெடுக்கப்படுகிறது, அவை படம் 5-இல் காட்டப்பட்டுள்ளது. எழுத்து வரி வெட்டுதல் வழிமுறை மற்ற ஆராய்ச்சியாளர்கள் கண்டறிந்த முறைகளை விட நல்ல விளைவைத் தருகின்றன. எழுத்து வரிசை வளைந்திருந்தாலும் அதனை சரியான எழுத்துக்களுடன் எழுத்துக்களின் வடிவம் மாறாமல் பிரிப்பது இதன் தனிச்சிறப்பாகும். எழுத்து வரி வெட்டுதல் முறையில் பிரிக்கப்பட்ட எழுத்து வரிசைகள் அடுத்த கட்ட செயலாக்கத்திற்கு உதவியாக அமைகிறது (Spurgen Ratheash, & Mohamed Sathik 2019).

படம் 5 எழுத்து வரி வெட்டுதல் முறையில் தனித்தனியாக பிரிக்கப்பட்ட எழுத்து வரிசைகள்

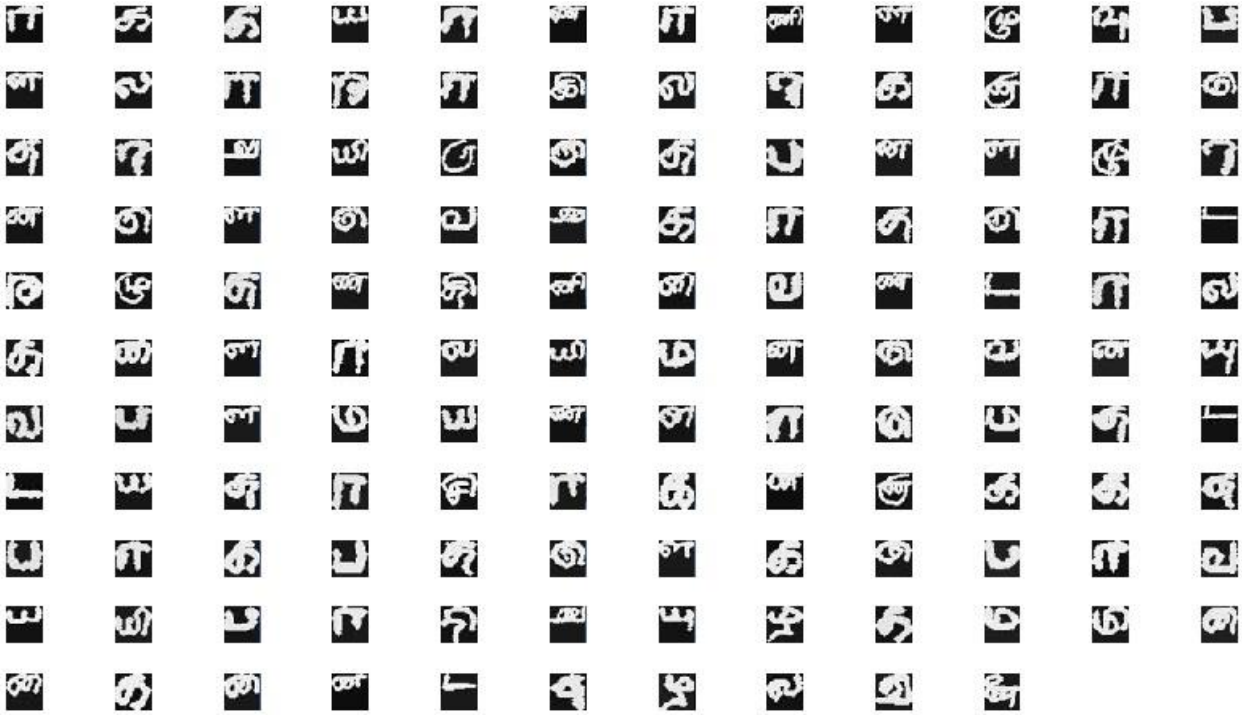
எழுத்துக்களைப் பிரித்தல்

பனை ஓலைச்சுவடிகளில் உள்ள தமிழ் எழுத்துக்கள் நீண்டு ஒரே எழுத்து வரிசையின் எழுத்துக்களுடன் ஒட்டி காணப்படும். இதனால் எழுத்துக்களை கணினி வழி அடையாளப் படுத்துவதற்கான தனி படங்கள் உருவாக்குதல் சிக்கலான நிலைக்குத் தள்ளப்படுகிறது. இத்தகைய சிக்கல்களை தீர்ப்பதற்காக ஆராய்ச்சியாளர்கள் கிடைமட்ட செங்குத்து வெட்டுதல் (Horizontal Vertical Slicing) வழிமுறையை பனை ஓலைச்சுவடிகளில் உள்ள தமிழ் எழுத்துக்களுக்காக பிரத்தியேகமாக உருவாக்கியுள்ளனர் (Mohamed Sathik, & Spurgen Ratheash, 2020). அதன்படி எழுத்து வரி வெட்டுதல் முறையில் தனியாக பிரிக்கப்பட்ட எழுத்து வரிசையில் உள்ள எழுத்துக்கள் ஒன்றையொன்று ஒட்டிக் கொண்டும், பிண்ணிக் கொண்டும் இருந்தால் அதனை எழுத்துக்களின் வடிவம் மாறாமல் பிரித்து தனி தனி படங்களாக மாற்றப்படுகிறது அவை படம் 6-இல் காட்டப்பட்டுள்ளது.

படம் 6 கிடைமட்ட செங்குத்து வெட்டுதல் முறையில் தனித்தனியாக பிரிக்கப்பட்ட எழுத்துக்கள்

எழுத்துக்களை அடையாளப்படுத்துதல்

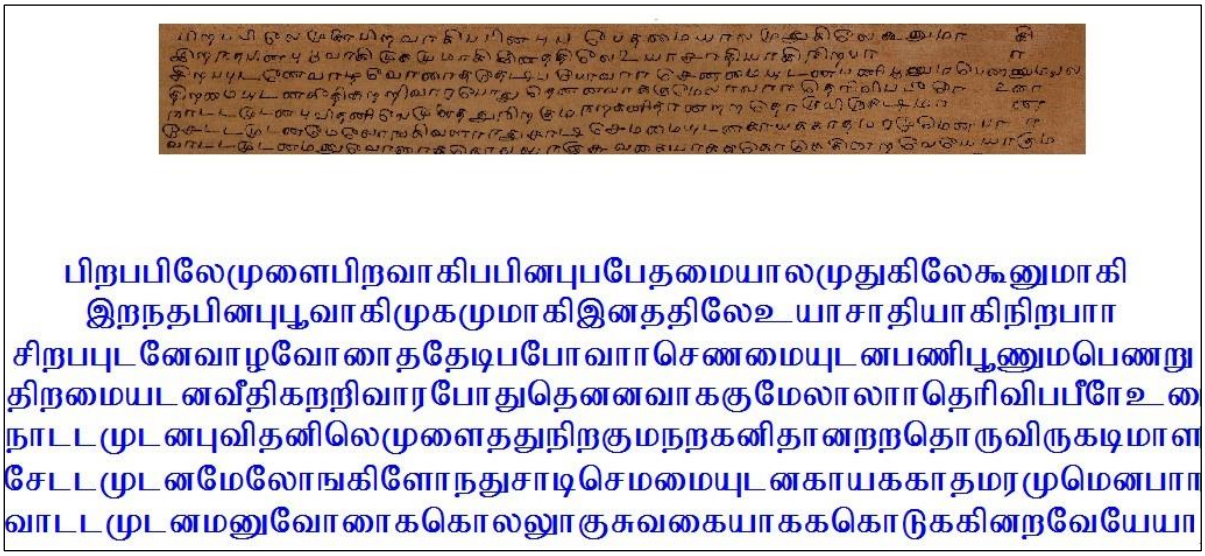
பனை ஓலைச்சுவடிகளில் உள்ள தமிழ் எழுத்துக்களை கணினி வழி அடையாளப்படுத்துதலில் தொடர்ச்சியான நரம்பணுப் பிணைய (Recurrent Neural Networks) நுட்பம் பயன்படுத்தப்படுகிறது. தமிழ் எழுத்துக்களை அடையாளப்படுத்துவதில் பல்வேறு நுட்பங்களை பல்வேறு ஆராய்ச்சியாளர்கள் பயன்படுத்தியிருந்தாலும் பனை ஓலைச் சுவடிகளின் தமிழ் எழுத்துக்களை அடையாளப்படுத்துவதில் பயன்படுத்தப்பட்டுள்ள தொடர்ச்சியான நரம்பணுப் பிணைய நுட்பம் அதிகபட்ச எழுத்துக்களை அடையாளப்படுத்துகிறது. எழுத்து வரிசை வெட்டுதல் முறைப்படி பிரிக்கப்பட்ட எழுத்து வரிசைகள் கிடைமட்ட செங்குத்து வெட்டுதல் முறைப்படி தனி தனி எழுத்து படங்களாக பிரித்த பிறகு தொடர்ச்சியான நரம்பணுப் பிணைய நுட்ப முறை பயன்படுத்தப்படுகிறது. இம்முறையை பயன்படுத்துவதற்கு ஒருசில முன் செயலாக்க முறைகள் பின்பற்றப்படுகிறது. அதன்படி கிடைமட்ட செங்குத்து வெட்டுதல் முறைப்படி பிரிக்கப்பட்ட எழுத்துக்கள் வெவ்வேறு அளவுகளைக் கொண்டுள்ளதாக இருக்குமாயின் முதலில் அவற்றை எல்லாம் இயல்பாக்குதல் (Normalization) முறை மூலம் ஒரே அளவாக மாற்றப்படுகிறது, அவை படம் 7-இல் காட்டப்பட்டுள்ளது.



படம் 7. பிரிக்கப்பட்ட எழுத்துக்கள் ஒரே அளவாக மாற்றப்படுதல்.

அதன்பின் நான்கில் இரண்டு பங்கு படங்கள் பயிற்றுவிப்பதற்கும் (Training) மற்றவை சோதனை (Testing) செய்வதற்கும் பிரிக்கப்படுகிறது. பயிற்றுவித்தல் என்பது எழுத்துக்களின் உருவங்களின் அம்சங்களை (features) கணினியை அதற்கேற்றவாறு புரிந்துகொள்ள வைப்பது. சோதனை என்பது கணினி புரிந்து கொண்ட எழுத்துக்களின் அம்சங்களை ஒத்துப்பார்த்து சரியான எழுத்தை கணிக்கச் செய்வது. இந்த ஆராய்ச்சியில் தொடர்ச்சியான நரம்பணுப் பிணைய நுட்பத்தின் இருதரப்பு நீண்ட கால

நினைவக (Bidirectional Long Short Term (Bi-LSTM)) வகைப்படுத்தி உபயோகப்படுத்தப்படுகிறது. இதில் நினைவகம் (memory cell) இருப்பதால் எழுத்துக்களின் அம்சங்களை நினைவில் வைத்து சோதனை எழுத்துக்களுடன் ஒப்பிடுவது எளிதாக்கப்படுகிறது (Mohamed Sathik, & Spurgen Ratheash 2020; Mohamed Sathik, & Spurgen Ratheash 2020A). இதனால் பல்வேறு காலக்கட்டத்தில் எழுதப்பட்ட தமிழ் எழுத்துக்கள் கணினி வழியாக மிக நேர்த்தியாக அடையாளப்படுத்தப்படுகிறது, அவை படம் 8-இல் காட்டப்பட்டுள்ளது.



படம் 8. கணினியால் அடையாளப்படுத்தப்பட்ட ஓலைச்சுவடியில் உள்ள தமிழ் எழுத்துக்கள்.

முடிவுரை

தமிழ் எழுத்துக்களை கணினி வழியாக அடையாளப்படுத்துவதில் இது வரை பேனாவால் எழுதப்பட்ட தமிழ் எழுத்துக்களே அதிகமாக செயல்முறைப்படுத்தப்பட்டுள்ளது. பனை ஓலைச் சுவடிகளில் உள்ள தமிழ் எழுத்துக்களை கணினி வழியாக அடையாளப்படுத்துவது மிக குறைந்த அளவிலேயே செயல்முறைப்படுத்தப்பட்டுள்ளது. அவ்வாறு செய்திருந்தாலும் மிக குறைந்த எண்ணிக்கையிலான எழுத்துக்களை அடையாளப்படுத்தியுள்ளது. இந்த ஆராய்ச்சியில் பனை ஓலைச் சுவடிகளில் உள்ள அனைத்து தமிழ் எழுத்துக்களையும் கணினிஅடையாளப்படுத்துவது இதன் சிறப்பாக கருதப்படுகிறது. இது 99.5% சரியான எழுத்துக்களை அடையாளப்படுத்துகிறது. இது மற்ற ஆராய்ச்சியை விட அதிக வெற்றி விகிதத்தில் நல்ல விளைவைக் கொடுக்கிறது. தமிழ் எழுத்துக்கள் பொறிக்கப்பட்ட கல்வெட்டுகளில் இந்த முறையை பயன்படுத்தி எழுத்துக்களை அடையாளப்படுத்துவதே இந்த ஆராய்ச்சியின் எதிர் கால திட்டம் ஆகும். தமிழ் எழுதுமுறையில் இரண்டு எழுத்துக்களை ஒரு எழுத்தில் மறைத்து வைத்திருக்கும் சிறப்பம்சத்தை எந்த சிக்கலுமின்றி துல்லியமாக கண்டறிந்து வெளிக்கொணர்வதே இந்த ஆராய்ச்சியின் மிக முக்கிய அம்சம் ஆகும்.

References

Banumathi, P., Nasira, G.M., (2011) Handwritten Tamil Character Recognition Using Artificial Neural Networks, 2011 International Conference on Process Automation, Control and Computing, 1-5. https://doi.org/10.1109/PACC.2011.5978989

Challa, N.P., & Mehta, R.V.K. (2017) Applications of Image Processing Techniques on Palm Leaf Manuscripts-A Survey. *Helix: The Scientific Explorer*, *7*(5), 2013-2017.

Deepa, M., Deepa, R., Meena, R., Nandhini, R., (2019) Tamil handwritten text recognition using convolutional neural networks, International Journal of Engineering Science and Computing, 9(3) 20986-20988.

Kannan R.J., Subramanian S., (2015) An Adaptive Approach of Tamil Character Recognition Using Deep Learning with Big Data-A Survey. In: Satapathy S., Govardhan A., Raju K., Mandal J. (eds) Emerging ICT for Bridging the Future - Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1. Advances in Intelligent Systems and Computing, vol 337. Springer, Cham. https://doi.org/10.1007/978-3-319-13728-5_63

Kattalai Kailasam, V., (2018) Yedum Eluthum, First Edition, Malli Padhipagam, Tirunelvlei, TamilNadu, India.

Kesiman, M.W.A., Valy, D., Burie, J.-C., Paulus, E., Suryani, M., Hadi, S., Verleysen, M., Chhun, S., Ogier, J.-M., (2018) Benchmarking of Document Image Analysis Tasks for Palm Leaf Manuscripts from Southeast Asia, Journal of Imaging, 4(2) 43. https://doi.org/10.3390/jimaging4020043

Mohamed Sathik, M., Spurgen Ratheash, R., (2020) Optimal Character Segmentation for Touching Characters in Tamil Language Palm Leaf Manuscripts using Horver Method, International Journal of Innovative Technology and Exploring Engineering, 9(6) 1010-1015. https://doi.org/10.35940/ijitee.E3126.049620

Mohamed Sathik, M., Spurgen Ratheash, R., (2020A) Handwritten Tamil Character Recognition in Palm Leaf Manuscripts using BiLSTM Classifier, International Journal of Future Generation Communication and Networking, 13(4) 4440-4450.

Sabeenian R.S., Paramasivam M.E., Anand R., Dinesh P.M., (2019) Palm-Leaf Manuscript Character Recognition and Classification Using Convolutional Neural Networks. In: Peng SL., Dey N., Bundele M. (eds) Computing and Network Sustainability. Lecture Notes in Networks and Systems, vol 75. Springer, Singapore. https://doi.org/10.1007/978-981-13-7150-9_42

Sornam M., Vishnu Priya C., (2018) Deep Convolutional Neural Network for Handwritten Tamil Character Recognition Using Principal Component Analysis. In: Bhattacharyya P., Sastry H., Marriboyina V., Sharma R. (eds) Smart and Innovative Trends in Next Generation Computing Technologies. NGCT 2017. Communications in Computer and Information Science, vol 827. Springer, Singapore. https://doi.org/10.1007/978-981-10-8657-1_61

Spurgen Ratheash, R., Mohamed Sathik, M., (2019) Line Segmentation Challenges in Tamil Language Palm Leaf Manuscripts, International Journal of Innovative Technology and Exploring Engineering, 9(1) 2363-2367. https://doi.org/10.35940/ijitee.L3159.119119

Vellingiriraj, E.K., Balasubramanie, P., (2014) Recognition of Ancient Tamil Handwritten Characters in Historical Documents by Boolean Matrix and BFS Graph, International Journal of Computer Science and Technology, 5(SPL – 1) 65-68.

Funding: NIL
Acknowledgement: NIL
Conflict of Interest: NIL
About the License:

Attribution 4.0 International (CC BY 4.0)

© The author 2021. The text of this article is licensed under a Creative Commons Attribution 4.0 International License