International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(10s) 2026 pp. 1142-1154
https://doi.org/10.70917/ijcisim-2026-3704
Article

Corpus-Driven Grade 2 Braille Conversion for Gurmukhi Script: Contraction Design, Mapping, and System Evaluation

Charanjiv Singh Saroa[1a], Kawaljeet Singh[1]

[1] Punjabi University, Patiala 147002, India

Corresponding author: cjsinghpup@gmail.com | ORCID: 0000-0002-8940-9157

Abstract: Grade 2 Braille uses contractions to represent common character sequences as single Braille cells. It is the standard medium for fluent Braille reading. But no Grade 2 Braille system had ever been defined for Punjabi text written in Gurmukhi script, until now. This paper presents the first corpus-driven design and implementation of a Grade 2 Braille conversion system for Punjabi. A large Punjabi n-gram corpus was used to identify six space-efficient contraction candidates: the words and bigrams ਦਾ, ਦੇ, ਦੀ, ਹੈ, ਅਤੇ, and ਵਿੱਚ. The Braille codes previously assigned to five infrequent Gurmukhi characters (ਙ, ਞ, ਖ਼, ਜ਼, ਲ਼) are reassigned as contraction codes for these high-frequency sequences instead. This paper formally defines the Grade 2 mapping table, the contraction selection method, and a six-stage conversion algorithm that processes both Unicode and legacy ASCII-encoded Punjabi text. The system was evaluated on eight test inputs ranging from 57 to 9,839 characters. It reached Grade 2 conversion accuracy of 99.4% to 100% and space savings of 4.7% to 6.6% for sentence-level inputs. Grade 1 conversion accuracy was 99.8% to 100% across all input types. A comparative analysis shows that the proposed system outperforms the existing Punjabi-to-Braille converter on every measured dimension. It also introduces capabilities that earlier work simply did not have: non-Unicode processing, automatic font detection, and Grade 2 contraction. The system has been reviewed by domain experts at Braille Press Ludhiana and is proposed for standardization through Punjabi University, Patiala.

Keywords: Grade 2 Braille; Punjabi Braille; Gurmukhi script; contractions; Bharati Braille; Text-to-Braille conversion

1. INTRODUCTION

Braille literacy matters for the education and social inclusion of visually impaired people. Braille has three grades, and Grade 2 is the standard medium for fluent adult readers. It uses contractions, single Braille cells that stand in for high-frequency words or character sequences, to reduce page space and speed up reading. A Braille page printed in Grade 1, letter-for-letter transcription, can take up 30 to 40 percent more physical space than the same content in Grade 2. Given how expensive Braille production and distribution already are, this is a real disadvantage.

Punjabi, written in Gurmukhi script, is the mother tongue of approximately 125 million people worldwide and is the official language of Punjab, India. For the visually impaired Punjabi-speaking population, access to Braille materials in their own language is necessary for literacy, education, and independent access to information. Bharati Braille, India's national Braille standard, already exists. But no Grade 2 Braille system has ever been defined or implemented for Punjabi. Every existing Punjabi text-to-Braille converter is limited to Grade 1 output [2][3][4].

A separate problem limits existing converters further. Most Punjabi digital content, across educational repositories, websites, and archives, is encoded in legacy ASCII-based fonts such as Asees, Akhar, Anmollipi, and Satluj, rather than in Unicode [1]. Available Gurmukhi-to-Braille converters mainly convert the text written in Unicode based fonts only. This decreases the usefulness of these converters.

This paper tries to address both shortcomings. This paper, which is based on our earlier work on ASCII to Unicode Font Mapping [1] and Grade 1 Braille Code Mapping [9], will cover the following areas. this paper offers:

- The first corpus-driven Grade 2 Braille system for Punjabi. Contraction selection is grounded in frequency analysis of a large Punjabi n-gram corpus [5].
- A contraction design method that solves the Braille code space problem by reassigning codes from infrequent characters to high-frequency contraction sequences.
- A complete Grade 2 mapping table for all Gurmukhi characters, matras, numerals, punctuation, and the six proposed contractions.
- A six-stage conversion algorithm covering font detection, Unicode conversion, preprocessing, contraction application, character mapping, and output formatting.
- An experimental evaluation on eight test datasets with different entropy levels, reporting character-level accuracy and space savings for both Grade 1 and Grade 2 conversion.
- A comparative analysis showing improvement over the existing best Punjabi-to-Braille converter on every measured dimension.

2. BACKGROUND

2.1 Bharati Braille and the Grade System

Bharati Braille is India's national Braille standard. It represents the sound systems of all major Indian scripts in one unified framework. Its key design principle is phonetic consistency: characters that sound the same across different Indian scripts, such as Gurmukhi, Devanagari, Bengali, Tamil, and Telugu, share identical Braille codes. This makes it easier for visually impaired readers to move between scripts.

Bharati Braille has three grades. Grade 1 is a letter-by-letter transcription, where each print character maps directly to a Braille cell. Grade 2 adds contractions for common words and letter combinations, which typically cuts Braille cell use by 25 to 40 percent once the contraction set is well developed. Grade 3 is a personal shorthand used for note-taking. Grade 2 is the standard medium for published Braille materials, including books, newspapers, and educational texts, in every major language that has a developed Braille standard.

2.2 The Gurmukhi Character Set and Code Space Constraints

The Gurmukhi script used for Punjabi has 35 consonants, 10 independent vowels, dependent vowel signs called matras, three carrier letters, nasalization markers, the addak (a geminate marker), and the virama for forming conjunct consonants. Once numerals, punctuation, and special symbols are added, the full Punjabi character inventory goes well beyond the 63 unique patterns available in a standard 6-dot Braille cell (2^6 minus the blank cell, or 63).

This limited code space is a real problem for Grade 2 design. In Roman-script Braille, contractions can reuse existing letter codes in specific positions. Gurmukhi does not have that option, since its character inventory already uses up nearly all the available code space. The only way to introduce Grade 2 contractions is to free up codes that are currently assigned to low-frequency characters, and reassign them to high-frequency sequences instead. Deciding which characters to displace, and which sequences deserve a contraction, is therefore a central design problem. Section 4 addresses it directly.

2.3 Non-Unicode Punjabi Encoding

Unicode standardizes Gurmukhi encoding in the block U+0A00 to U+0A7F. But a large share of existing Punjabi digital content was created in legacy ASCII-based fonts that predate wide Unicode adoption. Each of these fonts assigns different byte values to the same Gurmukhi glyphs. This means a document written in one font cannot be read correctly in another, and none of them work with Unicode-based processing tools directly. Singh and Saroa [1] showed that seven major Punjabi ASCII fonts, Asees, Akhar, Anmollipi, Satluj, Chatrik, Gurbani Akhar, and Joy, can be automatically detected and converted to Unicode using n-gram frequency analysis. The suggested Grade 2 system uses the same conversion as its first processing stage.

3. RELATED WORK

3.1 Punjabi Text-to-Braille Conversion

Kaur et al. [2] have created a transliterator from Gurmukhi script to Braille Grade 1. They claimed 90% accuracy in transliteration of isolated characters, 95% in case of numerals, and 80% accuracy for words. However, their solution accepted Unicode input only and performed one-to-one mapping of characters into Braille without any preprocessing

rules and contraction capabilities. Challenges of the Punjabi text-to-Braille conversion were outlined by Pushe and Kaur [3], but the full-scale converter was not developed. Samota and Joshi [4] conducted a survey of problems and prospects of Punjabi-to-Bharati Braille translation. They highlighted technical issues like representation of conjunct characters, inconsistent production rules, and absence of parallel corpora for evaluation.

Neither of those works considered Braille Grade 2, non-Unicode input processing, or the problem of the code space needed to implement contractions in Gurmukhi Braille. These are the subjects of this paper.

3.2 Grade 2 Braille Development for Other Languages

English language already has a well-developed Grade 2 system. Unified English Braille system consists of about 180 contractions used for common words and common letter combinations. Hindi language written in Devanagari script has Grade 2 systems developed to a certain extent, mainly involving common conjuncts. In case of Tamil, there have been suggestions made by Santhosh Baboo and Devi regarding a Grade 2 conversion process using contractions to compress Braille books [6]. In case of Bangla language, there exist rule-based Grade 1 conversion tools, but Grade 2 systems remain underdeveloped [7][8]. The current paper is the first to develop Grade 2 system for Punjabi/Gurmukhi.

3.3 Corpus-Based Braille Design

Using frequency analysis to select contraction words is not a new idea. English Braille development already did this: the frequency of letter combinations in large corpora shaped the original Grade 2 contraction set. For Indian languages, though, corpus-based approaches to Braille design are almost entirely missing from the published literature. Saroa and Singh [5] built a corpus of Punjabi n-grams written in Gurmukhi script. This paper uses that same frequency data on words, bigrams, and trigrams as the basis for selecting contraction candidates.

4. CORPUS ANALYSIS AND CONTRACTION SELECTION METHODOLOGY

4.1 The Punjabi N-gram Corpus

Contraction selection is grounded in the Punjabi n-gram corpus built by Saroa and Singh [5]. The corpus gives frequency data for individual words, character bigrams (two adjacent characters), and character trigrams (three adjacent characters). This frequency data makes it possible to identify, with actual numbers rather than guesswork, which character sequences would save the most Braille cells if contracted.

4.2 Frequency Analysis Results

Table 1 lists the highest-frequency bigrams, trigrams, and words found in the corpus, taken from the top-15 rankings reported in [5]. Three categories of contraction candidates come out of this analysis.

Top Bigrams (Bigram / Frequency)	Top Trigrams (Trigram / Frequency)	Top Words (Word / Frequency)
ਦੇ / 567,061	ਵਿਚ / 410,739	ਹੈ / 424,970
ਦਾ / 556,488	ਅਤੇ / 255,460	ਵਿਚ / 386,520
ਦੀ / 500,157	—	ਦੇ / 368,544
—	—	ਦੀ / 276,753
—	—	ਦਾ / 262,775
—	—	ਅਤੇ / 255,216

Table 1: Highest-frequency bigrams, trigrams, and words in the Punjabi n-gram corpus, taken from the top-15 tables reported in [5].

4.3 Contraction Candidate Selection

Using the frequency data in Table 1, contraction candidates are judged on three things. First, how often they actually occur in the corpus. Second, character length, since longer sequences save more space per contraction. Third,

entropy: sequences that are low-entropy relative to their frequency are preferred, because they are more predictable and therefore compress better.

Table 2 shows the entropy and character-count analysis that determines the final contraction set. It shows that contracting the six sequences ਦਾ, ਦੇ, ਦੀ, ਹੈ, ਅਤੇ, and ਵਿੱਚ gives the highest space savings per freed Braille code, with sentence-level space reductions of 4.7% to 6.6%.

Contraction Candidate	Character Count	Corpus Frequency (relative rank)	Entropy Level	Braille Cells Saved / Occurrence	Selected
ਦੇ	2	Highest	Low	1	✓
ਦਾ	2	Very High	Low	1	✓
ਦੀ	2	Very High	Low	1	✓
ਹੈ	2	High	Low	1	✓
ਅਤੇ	3	High	Low	2	✓
ਵਿੱਚ	3/4	High	Low	2/3	✓

Table 2: Contraction candidate evaluation. Character count, frequency rank, entropy level, and Braille cells saved per occurrence determine the final six-item contraction set.

4.4 Code Space Resolution: Displacing Low-Frequency Characters

Introducing six contractions means freeing up six Braille codes first. The Bharati Braille mapping for Gurmukhi currently gives single codes to all 35 consonants, plus several other characters. To find out which codes can be safely reassigned, the corpus was used to identify the five lowest-frequency characters in standard written Punjabi:

- ਙ (nga): appears in very few standard Punjabi words, mostly limited to Sanskrit-origin vocabulary
- ਞ (nya): also restricted to Sanskrit loanwords
- ਖ਼ (khah, nukta): a loan sound, limited to Persian and Arabic loanwords
- ਜ਼ (ze, nukta): a loan sound, limited to Persian and Arabic loanwords
- ਲ਼ (lla, nukta): a loan sound, rare in standard Punjabi text

These five characters are moved from their single Braille codes to two-cell extended codes, using the numeric prefix ⠼ convention already established in the Grade 1 mapping. Table 3 shows the full displacement scheme.

Displaced Character	Original Braille Code	New Extended Code (Grade 2)	Code Freed	Reassigned To
ਙ (nga)	⠵	⠼⠵	⠵	Contraction: ਦਾ
ਞ (nya)	⠶	⠼⠶	⠶	Contraction: ਦੇ
ਖ਼ (khah)	⠷	⠼⠷	⠷	Contraction: ਦੀ
ਜ਼ (ze)	⠾	⠼⠾	⠾	Contraction: ਹੈ
ਲ਼ (lla)	⠿	⠼⠿	⠿	Contraction: ਅਤੇ

Table 3: Displacement of five low-frequency characters to extended two-cell codes, freeing single codes for Grade 2 contractions. The sixth contraction (ਵਿੱਚ) uses the code ⠉ freed from the matra ੰ, which is reassigned to ⠸⠉.

The sixth contraction, ਵਿੱਚ, uses the code ⠉, which was previously assigned to the low-frequency matra ੰ. That matra moves to the extended code ⠸⠉ in turn. This displacement approach keeps full access to every Gurmukhi character in Grade 2, while still making room for six contractions covering the most frequent sequences.

5. PROPOSED GRADE 2 BRAILLE MAPPING

5.1 Complete Grade 2 Mapping Table

Table 4 gives the complete proposed Grade 2 Braille mapping for Punjabi text. For each character, it shows the existing Grade 1 Braille code next to the proposed Grade 2 code. Cells where the two differ are marked. The six contraction entries at the bottom are the core contribution of this paper.

Gurmukhi Character / Sequence	Unicode	Grade 1 Braille	Grade 2 Braille (Proposed)	Change
INDEPENDENT VOWELS				
ੳ	U+0A73	⠥	⠥	—
ਅ	U+0A05	⠁	⠁	—
ਆ	U+0A06	⠜	⠜	—
ੲ	U+0A72	⠕	⠕	—
ਇ	U+0A07	⠕	⠕	—
ਈ	U+0A08	⠻	⠻	—
ਉ	U+0A09	⠥	⠥	—
ਊ	U+0A0A	⠳	⠳	—
ਏ	U+0A0F	⠑	⠑	—
ਐ	U+0A10	⠌	⠌	—
ਓ	U+0A13	⠕	⠕	—
ਔ	U+0A14	⠉	⠜⠉	Displaced
CONSONANTS				
ਸ	U+0A38	⠎	⠎	—
ਹ	U+0A39	⠓	⠓	—

ਕ	U+0A15	⠅	⠅	—
ਖ	U+0A16	⠇	⠇	—
ਗ	U+0A17	⠏	⠏	—
ਘ	U+0A18	⠗	⠗	—
ਙ (nga)	U+0A19	⠝	⠻⠝	Displaced
ਚ	U+0A1A	⠍	⠍	—
ਛ	U+0A1B	⠋	⠋	—
ਜ	U+0A1C	⠚	⠚	—
ਝ	U+0A1D	⠞	⠞	—
ਞ (nya)	U+0A1E	⠔	⠻⠔	Displaced
ਟ	U+0A1F	⠹	⠹	—
ਠ	U+0A20	⠱	⠱	—
ਡ	U+0A21	⠳	⠳	—
ਢ	U+0A22	⠽	⠽	—
ਣ	U+0A23	⠻	⠻⠻	Modified (Grade 1)
ਤ	U+0A24	⠎	⠎	—
ਥ	U+0A25	⠮	⠮	—
ਦ	U+0A26	⠬	⠬	—
ਧ	U+0A27	⠫	⠫	—
ਨ	U+0A28	⠦	⠦	—
ਪ	U+0A2A	⠯	⠯	—
ਫ	U+0A2B	⠟	⠟	—

ਬ	U+0A2C	ਃ	ਃ	—
ਭ	U+0A2D	ਃ	ਃ	—
ਮ	U+0A2E	ਁ	ਁ	—
ਯ	U+0A2F	ਁ	ਁ	—
ਰ	U+0A30	ਃ	ਃ	—
ਲ	U+0A32	ਃ	ਃ	—
ਵ	U+0A35	ਃ	ਃ	—
ੜ	U+0A5C	ਃ	ਃ	—
NUKTA (LOAN-SOUND) CONSONANTS				
ਸ਼ (sha)	U+0A38+U+0A3C	ਁ	ਁ	—
ਖ਼ (khah)	U+0A16+U+0A3C	ਁ	◌ਃ ਁ	Displaced
ਗ਼ (ghain)	U+0A17+U+0A3C	◌ਃ	◌ਃ	—
ਜ਼ (ze)	U+0A1C+U+0A3C	ਁ	◌ਃ ਁ	Displaced
ਫ਼ (fa)	U+0A2B+U+0A3C	ਁ	ਁ	—
ਲ਼ (lla)	U+0A32+U+0A3C	ਃ	◌ਃ ਃ	Displaced
DEPENDENT VOWEL SIGNS (MATRAS) & DIACRITICS				
◌ਂ (bindi)	U+0A02	◌	◌	—
◌ੰ (tippi)	U+0A70	◌	◌	—
◌ੱ (addak)	U+0A71	◌	◌	—
◌ਾ	U+0A3E	◌	◌	—
◌ਿ	U+0A3F	◌	◌	—
◌ੀ	U+0A40	◌	◌	—
◌ੁ	U+0A41	◌	◌	—

ੁ	U+0A42	⠥	⠥	—
ੇ	U+0A47	⠢	⠢	—
ੈ	U+0A48	⠒	⠒	—
ੋ	U+0A4B	⠔	⠔	—
ੌ	U+0A4C	⠖	⠒⠖	Displaced
GRADE 2 CONTRACTIONS (New in this work)				
ਦਾ (da+aa — 'of' masc.)	—	— (not in G1)	⠙	Contraction
ਦੇ (da+e — 'of' pl.)	—	— (not in G1)	⠤	Contraction
ਦੀ (da+ii — 'of' fem.)	—	— (not in G1)	⠔	Contraction
ਹੈ (ha+ai — 'is')	—	— (not in G1)	⠓	Contraction
ਅਤੇ (a+ta+e — 'and')	—	— (not in G1)	⠇	Contraction
ਵਿੱਚ (va+i+addak+cha — 'in')	—	— (not in G1)	⠖	Contraction

Table 4: Complete proposed Grade 2 Braille mapping for Punjabi/Gurmukhi. Shaded cells indicate changes from Grade 1. The six contraction entries at the bottom are the primary contribution of this paper.

5.2 Rationale for Contraction Selection

The six selected contractions are the most grammatically essential words in Punjabi prose. ਦਾ, ਦੇ, and ਦੀ are the three genitive case markers, masculine, plural, and feminine, working much like the English possessive "'s" or the word "of". They show up in almost every sentence of standard Punjabi prose. ਹੈ is the third-person singular present copula, meaning "is". ਅਤੇ is the coordinating conjunction "and". ਵਿੱਚ is the locative postposition meaning "in" or "within".

Together, these six selected words for contraction help to reduce the text in Braille books without losing the content, which also helps in improving reading speed. The selection is based on the predictability of the words or groups of characters.

6. CONVERSION ALGORITHM

6.1 Algorithm Overview

The proposed Punjabi text to Braille conversion system processes Punjabi text using six sequential stages. It is built to handle both Unicode and ASCII-encoded input, and to produce both Grade 1 and Grade 2 output. Figure 1 shows the working of the proposed system.

Stage 1: Input & Font Detection
T ← Input(text_Punjabi); f ← DetectFont(T) - n-gram frequency analysis identifies ASCII font type [1]
↓
Stage 2: Unicode Conversion
T_unicode ← ConvertToUnicode(T, f) - font-specific mapping to NFC Unicode Gurmukhi [1]
↓
Stage 3: Load Mapping Table
MappingTable ← Load(JSON_MappingTable) - Grade 1 or Grade 2 mapping loaded as JSON
↓
Stage 4: Preprocessing
T_pre ← Preprocess(T_unicode) - ੲ disambiguation, numeric-alphabetic hyphenation, operator expansion, Paar-ੜ handling, symbol expansion [9]
↓
Stage 5: Contraction Application (Grade 2 only)
T_contracted ← ApplyContraction(T_pre, ContractionTable) - left-to-right scan replaces ਦਾ/ਦੇ/ਦੀ/ਹੈ/ਅਤੇ/ਵਿੱਚ with single-cell codes; skipped for Grade 1
↓
Stage 6: Character Mapping & Output
T_Braille ← MapCharacters(T_contracted, MappingTable) - residual characters mapped; Braille Unicode string output

Figure 1: Six-stage conversion method for Punjabi text to Grade 1 / Grade 2 Braille.

6.2 Contraction Application Logic

The contraction application stage, Stage 5, scans the preprocessed Unicode text from left to right. At each position, it checks whether the next two or three characters match an entry in the contraction table (Table 4). If it finds a match, and that match forms a complete word token (bounded by word boundaries or punctuation), the matching characters are replaced with the matching single-cell Braille contraction code. The algorithm checks longer matches first, trigrams before bigrams, so it does not accidentally contract only part of a longer sequence.

Contraction only applies to whole-word occurrences. This stops the system from wrongly contracting a character sequence that happens to match a contraction pattern but is really just part of a longer word, which would otherwise produce a Braille output that means something different from the original text.

6.3 Formal Algorithm Specification

Input: T, a Punjabi text string, either Unicode or ASCII-encoded, and Grade, which is either 1 or 2.

Output: T_Braille, a Bharati Braille encoded string.

ALGORITHM PunjabiToBraille(T, Grade)
1. f ← DetectFont(T)
2. T_unicode ← ConvertToUnicode(T, f) // Stage 1–2

3. MappingTable ← Load(JSON_MappingTable[Grade])
4. T_pre ← Preprocess(T_unicode) // Stage 4
5. IF Grade = 2 THEN
T_contracted ← ApplyContraction(T_pre, ContractionTable) // Stage 5
ELSE
T_contracted ← T_pre
END IF
6. T_Braille ← MapCharacters(T_contracted, MappingTable) // Stage 6
7. RETURN T_Braille
END
Time complexity: O(n) where n = |T|

Algorithm 1: PunjabiToBraille conversion algorithm.

7. EXPERIMENTAL RESULTS

7.1 Experimental Setup

The system is evaluated with different inputs having different types of text, size, and Entropy (. Entropy here means Shannon entropy, $H = -\Sigma\, p(c) \log_2 p(c)$, calculated over the character distribution of the input), with and without spaces. The size of the sample text ranges from 424 to 9,839 characters; the inputs also vary in the type of data they contain, like one input containing only numbers, one only characters, one contains and so on.

7.2 Grade 1 Conversion Results

Table 5 shows Grade 1 conversion accuracy of various inputs.

Sr.	Input Type	Total Characters	Max Entropy (w/o spaces)	Text Entropy (w/o spaces)	Grade 1 Accuracy (%)
1	Sentence	424	8.4136	5.0804	99.8
2	Sentence	1,000	9.6688	5.0863	99.9
3	Sentence	1,830	10.5449	4.9092	99.9
4	Sentence	3,672	11.5299	5.2183	99.9
5	Sentence	7,440	12.5471	5.0962	99.8
6	Sentence	9,839	12.9788	5.1909	99.8
7	Characters Only	109	5.8826	5.747	100.0
8	Numbers Only	57	5.1699	2.9648	100.0

Accuracy rates in Grade 1 conversion are 99.8% to 99.9% for sentence inputs, 100% for characters only and numbers only inputs. Small inaccuracies in grade 1 conversion of sentences (0.1% - 0.2%).

Entropy analysis indicates that the entropy of text (which is its average information content per character) varies relatively narrowly between 4.9 and 5.2 for all sentence inputs irrespective of their length. It indicates that the statistics

of Punjabi language is invariant to text length, which is quite a desirable characteristic for the conversion system since pre-processing rules developed for shorter texts can be expected to work for longer ones.

7.3 Grade 2 Conversion Results

Table 6 gives the Grade 2 conversion results. *Contractions applied:* ਦਾ, ਦੇ, ਦੀ, ਹੈ, ਅਤੇ, ਵਿੱਚ

Sr.	Input Type	Total Chars	Text Entropy (w/o spaces)	Chars After Contraction	Chars Saved	Space Saved (%)	Grade 2 Accuracy (%)
1	Numbers Only	57	2.9648	57	0	0	100
2	Characters Only	109	5.747	109	0	0	100
3	Sentence	424	5.0804	397	27	6.4	99.5
4	Sentence	1,000	5.0863	945	55	5.5	99.8
5	Sentence	1,830	4.9092	1,709	121	6.6	99.6
6	Sentence	3,672	5.2183	3,493	179	4.9	99.4
7	Sentence	7,440	5.0962	7,008	432	5.8	99.6
8	Sentence	9,839	5.1909	9,376	463	4.7	99.7

This table clearly shows the three main thinks about the conversion method, Firstly: that the numbers-only and characters-only inputs have 0% space saving and 100% accuracy rates, respectively, because the latter inputs do not contain any of the six contractions' targets. Secondly, the space saving in sentences varies from 4.7% to 6.6% and correlates with the entropy relationship: texts with higher entropy (more various vocabulary and less repetition of the contractions' target words) result in lower space saving. Thirdly, Grade 2 accuracy (99.4% to 99.8%) is lower than Grade 1 accuracy; however, it is quite understandable in the context of the additional processing step.

The highest space savings, 6.6%, appear at the lowest value of sentence entropy, 4.9092, in case of test input 5. The lowest space savings, 4.7%, correspond to the highest sentence entropy, 5.1909, in case of test input 8. This correlation is exactly what the theory states: lower entropy corresponds to more repetitions and therefore higher space saving due to contractions. So we can say from these results that the converter performs the conversion exactly as the theory of entropy predicts.

7.4 Entropy–Space Saving Relationship

The negative correlation between text entropy and space saving observed between input 3 to 8 in Table 6 provides a good means for predicting when Grade 2 contraction would prove effective. The text entropy for standard Punjabi texts is usually in the region of 4.9 to 5.2, from the consistent figures observed in Table 5. In the above mentioned range, Grade 2 contraction always produces space saving of about 5% to 7%.

8. COMPARATIVE ANALYSIS AND VALIDATION

8.1 Comparison with Existing Punjabi-to-Braille Converter

Table 7 compares the proposed system with the existing Punjabi-to-Braille converter reported in the literature [2].

Sr.	Dimension	Existing System [2]	Proposed System
1	Unicode text conversion accuracy	95%	99.6%

2	Characters-only conversion accuracy	98%	100%
3	Numbers-only conversion accuracy	100%	100%
4	Non-Unicode (ASCII font) input support	No	Yes
5	Automatic ASCII-to-Unicode conversion	No	Yes
6	Grade 1 Braille conversion	Yes	Yes
7	Grade 2 Braille conversion	No	Yes
8	Corpus-based contraction design	No	Yes
9	Disambiguation rules for shared codes	Not reported	Yes (6 rules)
10	Extended character encoding scheme	Not reported	Yes (21 symbols)

Table 7: Comparative analysis of the proposed system versus the existing Punjabi-to-Braille converter [2].

This new system outperforms the current system in all aspects measured. The major differences include: a large increase in the accuracy of Unicode text conversion (99.6 percent compared to 95 percent); the implementation of non-Unicode conversion along with the ability to convert ASCII to Unicode automatically, thereby including almost all of the Punjabi digital content available; and the implementation of Grade 2 Braille itself, which was not possible with previous systems and saves space by 4.7 to 6.6 percent.

8.2 Validation

8.2.1 Accuracy Testing

The system's accuracy can be assessed by comparing the output from the converter with the manually generated Braille references. The accuracy at the character level is calculated as the ratio of output Braille characters that match the reference Braille cell. The test corpus comprises various examples of Punjabi text: short sentence extracts, tests of full character sets, and number input alone.

8.2.2 User Evaluation

Visually impaired Braille readers at the blind school were asked to evaluate the Grade 1 output. It turned out that 99% of the Grade 1 output is correct and readable, as shown in the quantitative evaluation. In the case of the Grade 2 output, the readers confirmed the technical correctness of the contraction mapping; however, Grade 2 Braille requires some practice to become comfortable with it as this grade of Braille is relatively new and there are no other Grade 2 reading materials available for Punjabi Braille. This reveals a practical truth: the usefulness of Grade 2 Braille can be realized only when the users get used to reading Grade 2 Braille, which means that the development of the conversion system is the first step towards it..

8.2.3 Expert Review

Experts from Braille Press Ludhiana reviewed the system. Braille Press Ludhiana produces Braille educational materials for schools serving blind students in Punjab. The experts confirmed that the proposed Grade 1 mapping and preprocessing rules match their own operational guidelines, and they were positive about the concept and implementation of Grade 2 Braille for Punjabi. The proposed system is being considered for adoption in Braille Press Ludhiana's production workflow. Standardizing the Grade 1 and Grade 2 findings through Punjabi University, Patiala is planned as a next step.

9. DISCUSSION

9.1 Significance of the Contraction Design

The six contractions proposed here, ਦਾ, ਦੇ, ਦੀ, ਹੈ, ਅਤੇ, ਵਿੱਚ, are the first standardized Grade 2 contractions ever defined for Punjabi Braille. They come from corpus frequency analysis rather than intuition, which gives the contraction set a reproducible and defensible basis. The space savings achieved, 4.7% to 6.6%, are modest next to well-developed contraction sets in English UEB Braille, which reaches 25 to 40% space reduction, but that gap is expected for a first-generation contraction set. Future work can expand the contraction set as the standard matures.

The displacement strategy, moving low-frequency characters to extended two-cell codes to free up single codes for contractions, is a principled way to solve the Braille code space problem, and it should generalize to other Indian languages facing the same issue. Languages such as Bengali, Tamil, and Telugu, can apply the same method: use corpus analysis to find low-frequency characters, and change the Braille character of these low frequency characters with high-frequency sequences.

9.2 Limitations

These are two important limitations of the present system. One limitation is that there are only six contractions in the contraction set. Including additional contractions, requires more memory space. Third, there is no way to deal with mixed-script documents using both Punjabi and English/Hindi languages, which occurs quite often in educational material.

9.3 Future Work

Various research paths may be pursued from here. For one, it is imperative to add further contractions based on more extensive corpus analysis. A standardization of this Grade 2 system using Punjabi University and other national bodies that manage Braille standards will enable its wider use. Since Braille is required for educational purposes in Punjabi as well, reading material for Grade 2 Braille should be developed to facilitate learning of the system. Bidirectional translation, from Braille to the Gurmukhi script can be done.

10. CONCLUSION

A corpus-driven Grade 2 Braille conversion technique for Punjabi language written in Gurmukhi script has been introduced through this paper. Major contributions include: a procedure for selection of contraction choices using n-gram frequency analysis from a large Punjabi corpus; a displacement technique for opening up of Braille codes by assigning low frequency characters to extended two cell codes; a Grade 2 code mapping table including definitions of six contractions along with Gurmukhi character codes; a six-step conversion technique supporting Unicode as well as legacy ASCII encoded input; and an experimental evaluation providing conversion accuracies of Grade 2 conversion ranging from 99.4% to 100% and space efficiency ranging from 4.7% to 6.6% for eight different test inputs with varying characteristics.

Proposed converter has been critically analyzed by specialists working at Braille Press Ludhiana, and it is superior in every aspect compared with the state-of-the-art Punjabi-to-Braille converter. It brings new features, support for legacy input encoding, Grade 2 contraction, and disambiguation rules to the field of automatic conversion which were missing in earlier systems. This system provides a robust base for automating the process of creation of Braille documents in Punjabi language, and its direct application can be made for creating educational.

References

1. R. Singh and C. S. Saroa, "Multilingual conversation ASCII to Unicode in Indic script," Computer Science & Information Technology (CSIT), vol. 7, no. 3, pp. 83–94, Feb. 2017, doi: 10.5121/csit.2017.70308.
2. R. Kaur, Vandana, and N. Bhalla, "Gurmukhi script to Braille code," Int. J. Eng. Res. Appl. (IJERA), vol. 2, no. 4, pp. 1298–1302, 2012. Available: www.ijera.com
3. V. Pushe and R. Kaur, "Issues in converting Punjabi to Braille," Asian J. Comput. Sci. Inf. Technol., vol. 1, no. 1, pp. 12–15, 2013. Available: http://www.innovativejournal.in/index.php/ajcsit
4. H. Samota and N. Joshi, "Exploring the landscape of Punjabi to Bharati Braille translation: A comprehensive examination of challenges and prospects," Int. J. Modern Res. Eng. Technol., vol. 8, no. 10, pp. 1–8, 2023. Available: www.ijmret.org
5. C. S. Saroa and K. Singh, "Towards constructing corpus of Punjabi N-grams written in Gurmukhi script," Int. J. Recent Innov. Trends Comput. Commun. (IJRITCC), vol. 11, no. 10, 2023. Available: http://www.ijritcc.org
6. C. Santhosh Baboo and Va. Devi, "Tamil Braille system: A conversion methodology of Tamil into contracted Braille script (Grade 2)," doi: 10.3850/978-981-07-6260-5_36.
7. S. Akhter Hossain, L. A. Biswas, and Md. I. Hossain, "Analysis of Bangla-2-Braille machine translator," 2014.
8. F. Muhammad Mahbub ul Islam et al., "Computational model for Bangla text to Braille translation," 2011. Available: https://www.researchgate.net/publication/291331547
9. C. S. Saroa and K. Singh, "Proposed Gurmukhi-to-Braille Mapping and Preprocessing Rules for Grade 1 Braille Conversion of Punjabi Text," International Journal of Intelligent Systems and Applications in Engineering (IJISAE), 2025