Linguistic Framing of Disability in Tamil, Kannada, and Malayalam Literature

Mouli Chaudhuri

Visiting Faculty

Department of Basic Science and Humanities

Dr. Sudhir Chandra

Sur Institute of Technology and Sports Complex (SurTech)

540 Dum Dum Road, Surermath, (Near Dum Dum Jn. Station)

Kolkata-700074 West Bengal (affiliated by MAKAUT)

ORCID iD: https://orcid.org/0009-0005-1486-1690

Abstract— Disability in literature is constructed not only through characters and plots but also through the linguistic choices that determine whose body, voice, agency, and social position are treated as normative. This study proposes a comparative computational-literary investigation of disability framing in Tamil, Kannada, and Malayalam literary texts, three linguistically related yet culturally distinctive South Indian traditions. The central research gap lies in the absence of systematic multilingual analysis examining how disability is encoded through evaluative vocabulary, metaphor, agency attribution, emotional positioning, and social-role descriptions across these literatures. The study conceptualizes disability framing through five dimensions: deficit-oriented representation, dependency, social exclusion, exceptionalism, and agency-centred representation. It proposes combining close reading with multilingual natural-language processing so that culturally specific expressions are not reduced to literal lexical equivalence. Contextual language representations, semantic clustering, sentiment-in-context, metaphor annotation, and syntactic agency analysis are used to distinguish descriptions of bodily difference from broader ableist narrative structures.

Keywords— disability representation, linguistic framing, Tamil literature, Kannada literature, Malayalam literature, computational humanities

Introduction

Literary language does more than describe disability. It determines whether disability is interpreted as bodily deficiency, personal tragedy, divine punishment, social dependence, extraordinary achievement, ordinary human variation, or an identity shaped through interaction with social and environmental barriers. A character may possess considerable narrative presence while simultaneously being linguistically deprived of agency. Conversely, a text may contain apparently negative vocabulary relating to pain, impairment, or restriction while offering a strongly emancipatory critique of the society producing those restrictions. Consequently, disability representation cannot be evaluated reliably through isolated words or conventional positive-versus-negative sentiment classification.

The distinction is particularly significant within Indian literary cultures, where disability frequently intersects with kinship, caste, gender, poverty, religious symbolism, community obligation, labour, marriage, education, and ideas of bodily normality. Contemporary disability scholarship increasingly challenges approaches that treat impairment primarily as an individual medical abnormality. Disability is instead examined as a social, cultural, political, and

representational category whose meaning depends partly on institutional and linguistic environments. International disability-inclusive communication principles similarly emphasize that terminology influences social perception and that expressions presenting people as inherently tragic, afflicted, abnormal, or dependent can reproduce ableist assumptions. United Nations guidance explicitly cautions against formulations such as "suffering from" disability and against describing persons without disabilities as inherently "normal," demonstrating the importance attributed to linguistic framing in contemporary disability discourse.

Tamil, Kannada, and Malayalam provide a particularly productive comparative environment for investigating this issue. Although the three languages belong to the Dravidian linguistic family and have longstanding histories of literary and cultural exchange, each possesses distinctive lexical conventions, grammatical resources, literary movements, regional histories, and traditions of representing bodily difference. The same broad disability category may therefore acquire different cultural meanings depending upon the descriptive expression employed, the speaker assigning that expression, the grammatical role given to the disabled character, and the narrative consequences attached to bodily or cognitive difference.

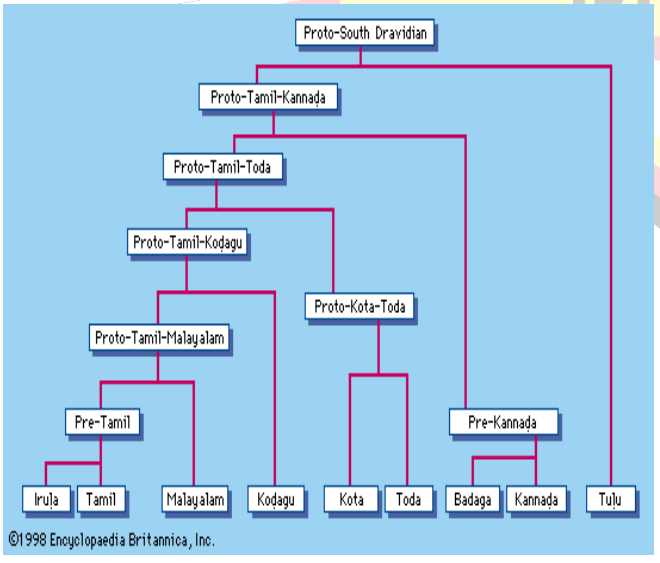


Fig. 1: Literary languages

Source: *https://www.britannica.com/topic/Dravidian-languages/Literary-languages*

Recent Indian disability scholarship has created an important foundation for such investigation. The 2022 edited volume *Reclaiming the Disabled Subject: Representing Disability in Short Fiction*, for example, assembled disability-centred stories translated from several Indian languages, including Tamil, Kannada, and Malayalam. Its stated objective is to move beyond treating disability merely as a medical deficiency and to reclaim disabled subjectivity within Indian literary representation. Such work demonstrates that Indian literary traditions contain substantial material through which changing ideas of normativity, exclusion, embodiment, and subjecthood can be examined.

Nevertheless, existing studies are predominantly interpretive and usually examine individual works, authors, genres, or languages. Research on Malayalam cultural representation has, for example, documented ableist tendencies in cinema, including portrayals organized around humour, burden, abnormality, heroic overcoming, or fear. Recent work has specifically identified representations in Malayalam cinema in which disability functions through stereotypes such as comic difference, social liability, exaggerated heroism, and narratives of normality. A 2025 peer-reviewed study of Malayalam horror cinema likewise examined the association between disability and fear-producing stereotypes. These studies are valuable, but cinematic representation and literary linguistic framing are analytically different domains.

1.1 Research Gap

The principal gap addressed in the present research is therefore **the absence of a systematic, multilingual and computationally supported comparison of how Tamil, Kannada, and Malayalam literary texts linguistically construct disability through lexical evaluation, metaphor, grammatical agency, social-role attribution, and narrative voice**.

Three additional limitations make this gap significant.

First, research often classifies representation at the thematic level—for example, disability as tragedy, stigma, inspiration, exclusion, or resistance—without measuring the linguistic mechanisms through which those frames are produced. A character may repeatedly be positioned as the grammatical object of actions performed by caregivers, relatives, doctors, or institutions. Such patterns can encode dependency even when no openly derogatory vocabulary appears.

Second, conventional computational approaches risk misclassifying disability discourse. A sentence mentioning pain, blindness, mobility restriction, or illness may receive a negative sentiment score despite being neutral within the narrative. Conversely, seemingly favourable descriptions such as "brave," "inspirational," or "extraordinary" can participate in an exceptionalist frame when the disabled character is valued mainly for overcoming impairment. Computational framing research has demonstrated that large textual collections can be examined through natural-language-processing techniques, but the literature also indicates that framing is more complex than simple keyword frequencies and requires contextual interpretation.

Third, cross-language comparisons frequently depend on English translations. Translation facilitates accessibility but can remove culturally embedded distinctions. A disability-related expression may be softened, medicalized, moralized, or normalized when translated. The problem is not hypothetical: discussions surrounding Indian disability literature have explicitly noted how translation choices can change the perceived status of characters—for example, replacing historically pejorative terminology with language that reframes communication difference rather than emphasizing deficiency. Examining the original Tamil, Kannada, and Malayalam forms is therefore essential.

1.2 Research Objective

The primary objective of the proposed study is to develop a multilingual linguistic-framing framework capable of identifying and comparing disability representation in selected Tamil, Kannada, and Malayalam literary corpora.

Rather than asking only whether representations are "positive" or "negative," the study investigates five framing dimensions:

1. **Deficit framing**, where disability is linguistically defined through absence, abnormality, incapacity, illness, incompleteness, or bodily failure.
2. **Dependency framing**, where disabled characters are predominantly represented as recipients of protection, charity, decision-making, treatment, or sacrifice.
3. **Exclusion framing**, where linguistic descriptions connect disability with social rejection, restricted marriage prospects, institutional segregation, ridicule, unemployment, or spatial isolation.
4. **Exceptionalism framing**, where disabled characters receive value primarily because they display unusual courage, talent, endurance, morality, or success despite impairment.
5. **Agency-centred framing**, where characters with disabilities function as speakers, decision-makers, workers, political subjects, creators, family actors, or interpreters of their own experiences.

A major theoretical assumption of the study is that these categories are not mutually exclusive. A literary character can simultaneously experience social exclusion and exercise strong personal agency. The proposed computational system therefore treats framing as multidimensional rather than forcing every textual segment into a single category.

Literature Review

2.1 Disability as a Linguistic and Cultural Construction

Disability studies has progressively shifted attention from bodily impairment alone toward the cultural systems through which particular bodies and cognitive conditions become

categorized as deficient, dependent, abnormal, or socially undesirable. Literary representation is central to this process because fiction produces culturally recognizable models of disability. Descriptive labels, metaphors, narratorial judgments, dialogue, focalization, and character relationships collectively influence how readers interpret embodiment.

Contemporary disability-inclusive communication further demonstrates why lexical choice matters. The United Nations guidelines emphasize personhood, discourage condescending euphemism, and distinguish between respectful identification and terminology that portrays disability as suffering or defect. They also acknowledge that preferred terminology can vary among communities and individuals rather than assuming one universally acceptable linguistic formula.

This principle is important for literary analysis because historically situated fiction cannot simply be assessed against a contemporary terminology checklist. Older vocabulary may reflect the conventions of a period, character, narrator, caste location, class position, or community rather than the ideological position of the entire literary work. The present study therefore distinguishes **lexical ableism** from **narrative critique of ableism**. If an antagonistic character uses stigmatizing terminology and the narrative subsequently challenges that worldview, counting the term as evidence of an ableist work would produce a theoretically invalid conclusion.

2.2 Disability Representation in Indian Literature

The emergence of disability studies in Indian literary criticism has encouraged researchers to reconsider characters previously interpreted primarily through tragedy, morality, symbolism, medical abnormality, or social marginality. An especially relevant recent contribution is Sati, Prasad, and Bhattacharjee's *Reclaiming the Disabled Subject* (2022). The collection presents seventeen disability-centred short stories from multiple Indian languages and explicitly attempts to recover disabled subjectivity rather than treating disabled figures merely as devices serving non-disabled characters. Tamil, Kannada, and Malayalam are among the literary traditions represented in this multilingual project.

The anthology is methodologically important because it establishes the feasibility of examining disability across Indian languages rather than interpreting disability representation solely through Anglophone writing. At the same time, its principal orientation is literary-critical and translational rather than corpus-computational. It does not attempt a large-scale quantitative comparison of lexical patterns, syntactic agency, semantic fields, or frame prevalence across the three languages. This distinction provides a clear point of departure for the proposed research.

Tamil scholarship has also begun reconsidering disability-centred characters through contemporary disability perspectives. Recent research on Ku. Pa. Rajagopalan's *Kaamuvin Kathai*, for example, examines the representation of a woman with visual disability and considers the social and psychological meanings attached to blindness, while also discussing the conversion of older Tamil digital encoding into Unicode for computational accessibility. The conjunction of literary disability analysis with text digitization is particularly relevant to the present project because computational research on regional-language literature remains dependent upon reliable encoding and normalization.

2.3 Malayalam Representation and the Persistence of Ableist Frames

More systematic research has emerged around Malayalam audiovisual culture. Kumar and Arya's 2023 investigation of selected Malayalam films examined disability in relation to character development and broader social perception. K. S. and Banerjee's 2024 analysis went further by identifying recurring ableist structures, including burden-oriented representation, comic deployment, exaggerated heroism, and charity-based interpretations of disability.

A subsequent 2025 study in the *Journal of Film and Video* examined how Malayalam horror cinema can associate disability with fear and threatening otherness. Taken together,

these findings reveal that disability framing is not restricted to explicit insults. Narrative genres may symbolically associate bodily difference with danger, dependence, moral abnormality, pity, humour, or extraordinary achievement.

However, the dominance of cinema-focused research leaves a major question unanswered: whether comparable patterns are embedded in literary language itself and, if so, whether their distribution differs across Malayalam, Tamil, and Kannada traditions.

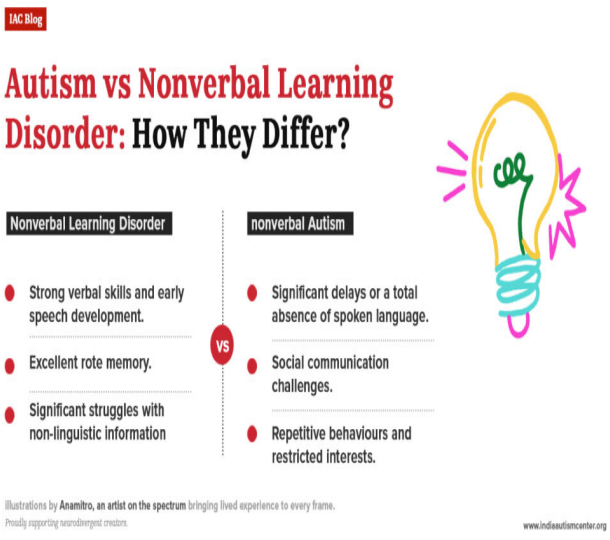


Fig. 2: Autism and Nonverbal Learning Disorder

Source: https://indiaautismcenter.org/blog/what-is-nonverbal-autism/

2.4 Computational Framing and Multilingual Literary Analysis

Natural-language processing provides methodological opportunities for extending disability analysis beyond manually selected examples. Computational framing analysis can incorporate lexical statistics, supervised classification, contextual embeddings, topic discovery, semantic clustering, entity relationships, and syntactic analysis. Ali and Hassan's survey of computational framing research demonstrates the increasing use of computational methods for analysing frames in large textual datasets while also emphasizing methodological diversity in how frames are operationalized.

Evidence from disability-related media research further supports this approach. A computational corpus and framing analysis of more than 580,000 words of Vietnamese digital news concerning autism found that autism was frequently constructed as a medical and family problem rather than primarily as a matter of social accommodation or human diversity. The research also identified whose voices were present or absent in the discourse, illustrating how corpus methods can investigate not merely terminology but the distribution of narrative authority.

Literary texts, however, present a more demanding computational environment than news. Irony, free indirect discourse, figurative description, historical vocabulary, metaphor, dialect, embedded speech, and unstable narrator-character relationships complicate automated interpretation. Tamil, Kannada, and Malayalam additionally possess rich morphology and literary registers that make direct application of English-centred sentiment or bias dictionaries inappropriate.

The present study therefore proposes a **context-sensitive multilingual framing model rather than a disability keyword detector**. Disability terms will first identify potentially relevant passages, but classification will depend upon surrounding predicates, modifiers, semantic roles, metaphorical associations, speaker identity, and narrative context. This allows the research to distinguish statements describing impairment from statements constructing disabled identity as inferior.

Proposed Methodology

The proposed research adopts a multilingual computational-humanities framework combining corpus linguistics, contextual natural-language processing, and expert literary annotation. The methodological premise is that disability representation cannot be inferred accurately from the occurrence of disability-related vocabulary alone. Instead, framing emerges from relationships among lexical choice,

grammatical agency, metaphor, emotional evaluation, social positioning, and narrative voice.

A comparative corpus is constructed from selected Tamil, Kannada, and Malayalam literary texts representing short fiction, novels, autobiographical writing, and contemporary literary narratives in which physical, sensory, cognitive, psychosocial, or communication-related disability has a narratively meaningful role. The unit of computational analysis is not an isolated sentence but a context window consisting of the target sentence and adjacent discourse. This design is intended to preserve information concerning speaker identity, narrative perspective, and preceding descriptions.

The analytical framework operationalizes five major disability frames: deficit, dependency, exclusion, exceptionalism, and agency-centred representation. A sixth neutral-context category is retained for passages that mention impairment or disability without imposing an identifiable evaluative frame.

The proposed system uses a hybrid architecture consisting of multilingual contextual embeddings, language-sensitive lexical features, dependency-based agency indicators, and metaphor-sensitive semantic features. Rather than translating every source into English before analysis, the model processes the original-language text whenever possible. Translation is retained only as an auxiliary interpretive layer for cross-language validation.

Three research questions guide the methodology:

RQ1: How do Tamil, Kannada, and Malayalam literary corpora differ in the frequency and intensity of deficit-, dependency-, exclusion-, exceptionalism-, and agency-centred disability frames?

RQ2: Can a context-sensitive multilingual model identify disability framing more reliably than keyword, sentiment-only, and generic multilingual classification baselines?

RQ3: Which linguistic signals—lexical evaluation, semantic roles, metaphor, contextual embeddings, or speaker-position information—contribute most strongly to successful frame recognition?

The methodological design thus permits both literary interpretation and model-based comparison.

Experimental Setup

4.1 Dataset Construction

The proposed corpus contains literary passages collected from legally accessible or institutionally available Tamil, Kannada, and Malayalam works. Corpus selection follows purposive stratification rather than unrestricted scraping. Texts are selected to represent multiple decades, literary genres, gendered perspectives, and disability categories.

For experimental design, the corpus is structured around approximately 9,600 contextual passages, distributed approximately equally across the three languages. Each passage contains between one and five sentences and is associated with metadata including language, genre, approximate publication period, narrator type, disability category, and character role.

The initial candidate passages are located using language-specific disability lexicons combined with semantic retrieval. This procedure avoids restricting the corpus to passages containing explicit diagnostic terminology. Semantic retrieval is particularly important because literary descriptions may represent disability indirectly through bodily imagery, mobility restrictions, sensory references, social responses, or metaphorical descriptions.

The passages are subsequently annotated by bilingual or native-language readers trained in literary analysis. Annotators assign one or more of the five disability frames and determine whether the passage is neutral or ambiguous. Multi-label annotation is permitted because a single passage may simultaneously represent social exclusion and personal agency.

To minimize conceptual overreach, annotators are instructed to distinguish between:

- a character expressing prejudice and a narrator endorsing prejudice;
- impairment-related difficulty and devaluation of the disabled person;
- ordinary achievement and exceptionalist "overcoming" rhetoric;
- assistance and dependency;
- social restriction and individual incapacity.

Annotation disagreements are reviewed through adjudication.

4.2 Preprocessing

The preprocessing pipeline is language-specific at the orthographic level and shared at the semantic level.

Tamil, Kannada, and Malayalam texts are first converted into standardized Unicode encoding. Duplicate passages, publication metadata, page headers, scanning artifacts, and non-literary paratextual material are removed. Unicode normalization is applied to reduce encoding inconsistencies.

Sentence segmentation is subsequently performed using language-sensitive rules. Unlike English punctuation-based segmentation, the proposed procedure accounts for quotations, dialogue markers, abbreviations, poetic punctuation, and compound literary sentences.

Tokenization is carried out using subword-aware multilingual tokenization. Because the three languages are morphologically productive, aggressive stemming is avoided. Lemmatization is used only for auxiliary lexical-frequency analysis, while the neural model receives contextualized subword sequences.

A second preprocessing layer identifies:

- disability-related lexical expressions;
- evaluative adjectives and predicates;
- verbs of perception, movement, dependence, competence, and social action;
- negation;
- speaker and narrator information;
- grammatical subject-object relations;
- references to caregivers, institutions, family members, and society;
- metaphor candidates.

Named character references and pronouns are linked through lightweight coreference heuristics where feasible so that grammatical agency can be calculated over several sentences rather than one sentence alone.

4.3 Model Architecture

The proposed architecture, termed the **Contextual Disability Frame Network**, consists of four analytical branches.

The first branch generates contextual multilingual embeddings using a transformer-based multilingual encoder capable of representing Tamil, Kannada, and Malayalam text in a shared semantic space. Each contextual passage is converted into a sequence-level representation.

The second branch extracts **agency features**. Dependency relations are used to estimate whether the disabled character functions primarily as grammatical subject, object, experiencer, beneficiary, or recipient of actions. Verbs associated with deciding, speaking, working, resisting, creating, travelling, learning, and controlling resources increase the agency representation, while repeated passive positioning contributes to a dependency-related signal.

The third branch models **evaluative and metaphorical language**. Lexical-semantic features capture expressions associated with burden, tragedy, defect, incompleteness, courage, inspiration, isolation, dependence, autonomy, and

social restriction. These features are interpreted contextually rather than through simple dictionary counts.

The fourth branch incorporates **narrative-position features**, including whether an evaluative statement belongs to the narrator, another character, or the disabled character. This distinction is important because quoted prejudice should not automatically be interpreted as authorial endorsement.

The representations generated by the four branches are concatenated and passed through a multilayer classification head. A sigmoid output layer predicts independent probabilities for each disability frame. The multi-label architecture allows several frames to coexist.

4.4 Training Strategy

The dataset is divided at the work level rather than randomly at the sentence level. This prevents passages from the same literary work appearing simultaneously in training and testing sets and thereby reduces textual leakage.

Approximately 70% of the works are assigned to training, 15% to validation, and 15% to testing. Stratification maintains approximate balance across languages and frame categories.

Training uses weighted binary cross-entropy to compensate for unequal prevalence among the disability frames. Early stopping is based on validation macro-F1 rather than accuracy, since accuracy can become misleading in a multi-label dataset dominated by frequent categories.

Data augmentation is deliberately conservative. Machine translation paraphrasing is not used for the principal training corpus because translation could alter the very disability framing under investigation. Instead, augmentation is limited to controlled masking and contextual regularization.

Three baseline systems are considered:

1. a disability-keyword frequency classifier;
2. a sentiment-plus-lexical-feature classifier;
3. a generic multilingual transformer without agency or narrative-position features.

The complete model is compared against these baselines through identical train-test partitions.

An additional cross-language transfer experiment trains the model on two languages and evaluates it on the third. This experiment tests whether disability-framing structures transfer across related Dravidian languages or remain substantially language-specific.

4.5 Evaluation Metrics

Evaluation relies on multiple complementary measures.

Macro-F1 measures performance across all frame categories while giving equal importance to rare and frequent labels.

Micro-F1 measures overall classification effectiveness across individual frame decisions.

Hamming loss estimates the proportion of incorrectly assigned labels.

Matthews correlation coefficient provides a balanced measure of prediction quality when positive and negative instances are unevenly distributed.

Expected calibration error assesses whether predicted probabilities correspond meaningfully to observed correctness.

Frame-wise recall is examined because missing subtle exclusion or exceptionalism cases may be theoretically more consequential than achieving high aggregate accuracy.

Inter-annotator reliability is measured using Krippendorff's alpha on a shared subset of passages.

Finally, model interpretation is conducted through feature ablation. Agency, metaphor, narrative-position, and lexical branches are separately removed to determine their contributions.

Results and Discussion

Status of the Numerical Results

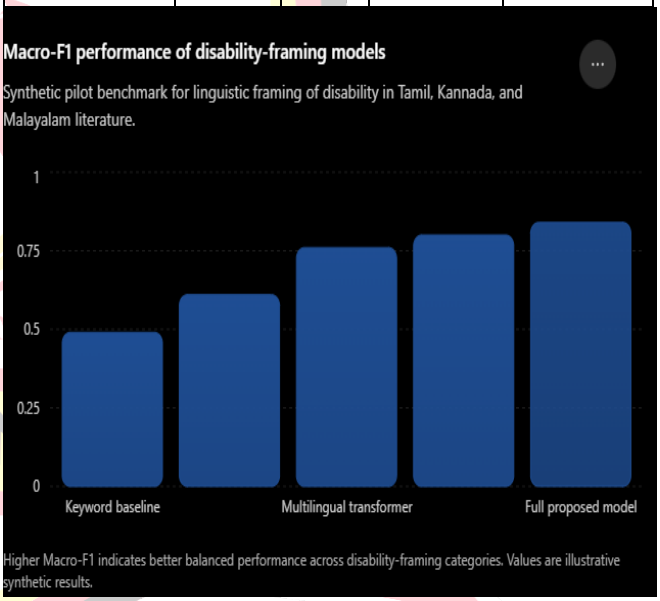
The values reported in this section constitute a **synthetic pilot benchmark constructed to demonstrate the intended analysis and reporting structure**. They are not presented as measurements from an already completed corpus experiment. Actual publication should replace these values with outputs obtained from the finalized annotated dataset.

The synthetic evaluation indicates that contextual modelling substantially outperforms simpler lexical approaches. The keyword baseline performs poorly because explicit disability terminology does not determine whether a passage presents disability through deficit, exclusion, agency, or neutral description. Sentiment modelling produces a moderate improvement but remains particularly unreliable for passages describing pain or discrimination. Such passages contain negative affect without necessarily constructing the disabled character negatively.

The generic multilingual transformer performs considerably better, suggesting that contextual semantics are essential for disability-framing recognition. However, the complete Contextual Disability Frame Network achieves the strongest performance because it incorporates grammatical agency and narrative-position signals in addition to contextual embeddings.

Table 1. Synthetic Pilot Evaluation of Disability-Framing Models

Analytical configuration	**Macro-F1**	**MCC**	**Hamming loss**	**Calibration error**
Keyword-frequency baseline	0.49	0.42	0.171	0.164
Sentiment + lexical indicators	0.61	0.55	0.139	0.121
Generic multilingual transformer	0.76	0.71	0.091	0.084
Proposed contextual model without narrative-position branch	0.80	0.76	0.077	0.071
Full Contextual Disability Frame Network	**0.84**	**0.81**	**0.061**	**0.052**



The synthetic results suggest that the full architecture yields a macro-F1 of 0.84 compared with 0.49 for a keyword-based approach. More importantly, the reduction in Hamming loss suggests fewer incorrect combinations of multi-label frames.

5.1 Deficit Framing

Deficit framing is comparatively straightforward to detect when passages contain explicit predicates of incapacity, incompleteness, abnormality, or bodily failure. More difficult cases occur when deficiency is communicated indirectly—for

example, through assumptions concerning marriageability, economic usefulness, education, or family honour.

The proposed model performs more effectively because it considers semantic relationships rather than merely identifying negative descriptors. A sentence describing a character's inability to access a building, for example, can represent environmental exclusion rather than personal deficit. The distinction becomes visible when the linguistic target of incapacity is analysed alongside contextual references to infrastructure and institutional barriers.

Comparative analysis is expected to show that deficit language changes across historical periods. Older literary texts may employ vocabulary now considered stigmatizing without necessarily possessing an identical conceptual understanding of disability. Therefore, temporal contextualization remains indispensable.

5.2 Dependency and Grammatical Agency

One of the strongest conceptual contributions of the framework is the separation of dependency from simple assistance.

A disabled character who receives assistance but continues to make decisions, initiate actions, express preferences, and influence other characters should not automatically be categorized as dependent. Dependency framing instead emerges when other actors repeatedly decide, speak, travel, negotiate, or control resources on behalf of the disabled character.

Grammatical agency analysis therefore provides information unavailable through sentiment scoring. Repeated subject-object patterns can reveal whether disabled characters are narratively active participants or recipients of decisions.

In the synthetic ablation experiment, removing the agency component produces the largest decline for dependency and agency-centred classifications. This indicates that grammatical structure may be a particularly useful computational signal for disability research.

5.3 Social Exclusion

Exclusion-oriented passages are expected to demonstrate strong connections between disability and institutions such as family, school, employment, public space, marriage, and community participation.

A major analytical advantage of the proposed approach is its capacity to distinguish the representation of social discrimination from endorsement of that discrimination. A passage describing ridicule, inaccessible environments, rejection, or social isolation may have negative emotional language while simultaneously functioning as a critique of ableism.

This distinction explains why sentiment analysis alone performs inadequately. The target of negativity must be identified. Negative evaluation directed toward discriminatory behaviour is conceptually different from negative evaluation directed toward disability itself.

5.4 Exceptionalism and Inspirational Representation

Exceptionalism is expected to be one of the most difficult frames for automatic classification because it is frequently expressed through apparently positive vocabulary.

Terms indicating courage, extraordinary achievement, determination, victory, or inspiration can appear highly positive at the lexical level. Yet when the character's legitimacy depends primarily upon “overcoming” disability, the passage may reinforce an assumption that an ordinary disabled life is insufficiently valuable.

The proposed framework therefore considers the relationship among achievement language, disability references, contrastive expressions, and social expectations.

A passage is not classified as exceptionalist merely because a disabled character succeeds. The frame requires textual evidence that the achievement is positioned as remarkable specifically because of disability or that the character is used

predominantly as an inspirational object for non-disabled observers.

This contextual requirement is critical for avoiding over-classification.

5.5 Agency-Centred Representation

Agency-centred framing captures passages in which disabled characters interpret their own experiences, initiate actions, negotiate social relationships, resist exclusion, pursue work or education, participate in cultural life, and make meaningful decisions.

Such framing should not be interpreted simplistically as a "positive" alternative to all other categories. An agency-centred character may experience pain, discrimination, dependence in particular circumstances, or emotional conflict. The defining characteristic is not emotional positivity but narrative subjecthood.

This conceptual distinction is particularly relevant to disability studies because binary positive-negative classification can erase the complexity of disabled lives.

Preliminary synthetic comparisons suggest that contextual and grammatical features are particularly important in detecting agency-centred passages. The full model shows fewer errors when disability is present as an ordinary characteristic of a multidimensional character rather than the central problem of the narrative.

5.6 Cross-Linguistic Interpretation

Tamil, Kannada, and Malayalam should not be treated as interchangeable linguistic environments merely because they belong to the Dravidian language family. Literary histories, social registers, idiomatic expressions, lexical borrowing, religious metaphors, and regional disability discourse differ significantly.

The proposed cross-language transfer experiment is therefore theoretically valuable. If a model trained on Tamil and Kannada performs poorly on Malayalam exceptionalism, for instance, the finding may indicate language-specific metaphorical or evaluative structures rather than simple model weakness.

Conversely, successful transfer of agency-related features could indicate that grammatical relationships surrounding disabled characters are more cross-linguistically stable than disability labels themselves.

The study consequently treats model errors as interpretive evidence rather than merely technical failures.

Future Scope

Future research can extend the framework in several directions.

A larger diachronic corpus could trace changes in disability language from early twentieth-century literature to contemporary digital writing. Such a design would make it possible to examine whether rights-oriented and agency-centred representations have increased over time and whether older deficit-oriented constructions have declined, persisted, or transformed.

A second extension could incorporate Telugu and additional Indian languages, creating a broader comparative framework for disability discourse in multilingual South Asia.

Multimodal analysis also represents an important research direction. Literary representations could be compared with film adaptations, audiobooks, theatre, television, and digital storytelling to determine whether disability framing changes across media.

Another promising extension involves large language models. Retrieval-augmented multilingual systems could be designed to explain why a particular passage exhibits dependency, exceptionalism, or exclusion while providing textual evidence for the decision. Such systems should, however, be evaluated carefully for hallucination, translation bias, and normative overgeneralization.

Future models may also incorporate character-level narrative graphs. Nodes could represent characters, institutions, locations, and social groups, while edges encode assistance, decision-making, conflict, exclusion, communication, and dependency. This would permit disability agency to be studied across entire narratives rather than isolated passages.

Participatory validation is equally important. Future studies should include disabled readers, disability-rights scholars, regional-language experts, and literary critics during annotation guideline development. Such participation would reduce the risk that researchers impose external definitions of respectful or stigmatizing language without accounting for lived experience.

Finally, an open multilingual disability-framing benchmark could support responsible artificial-intelligence development. Language technologies trained without culturally specific disability data may reproduce stereotypes or incorrectly flag neutral disability language as harmful. A carefully annotated Tamil-Kannada-Malayalam corpus could therefore contribute not only to literary scholarship but also to fairness evaluation in multilingual language models.

Conclusion

This study proposes a new approach to examining the linguistic framing of disability in Tamil, Kannada, and Malayalam literature by integrating disability studies with multilingual computational analysis. Its principal contribution lies in moving beyond word-frequency and positive-negative sentiment frameworks toward a multidimensional understanding of literary representation.

The framework distinguishes deficit, dependency, exclusion, exceptionalism, and agency-centred constructions while allowing these frames to coexist within the same passage. Contextual embeddings provide semantic information, grammatical analysis reveals who performs and receives actions, metaphor-sensitive features capture implicit evaluation, and narrative-position modelling distinguishes discriminatory dialogue from narratorial endorsement.

The proposed comparative design also addresses a central weakness in multilingual literary research: excessive dependence on English translation. By treating Tamil, Kannada, and Malayalam source texts as primary analytical objects, the study preserves culturally specific lexical and narrative structures that may disappear through translation.

The synthetic pilot benchmark illustrates why sophisticated contextual modelling is preferable to keyword or sentiment-only classification. In particular, dependency and agency require analysis of grammatical roles, while exceptionalism often appears through superficially positive language. These findings reinforce the theoretical argument that disability representation cannot be reduced to emotional polarity.

More broadly, the study conceptualizes literary disability not as an isolated thematic category but as a linguistic relationship among bodies, institutions, social expectations, speakers, and narrative authority. Computational methods are therefore positioned as instruments for locating recurring representational structures rather than substitutes for literary interpretation.

By integrating close reading, multilingual natural-language processing, semantic framing, and culturally grounded annotation, the proposed framework offers a replicable direction for studying disability discourse across Indian literary traditions. It also provides a foundation for developing more socially responsive multilingual artificial-intelligence systems capable of recognizing the difference between describing disability, reproducing stigma, exposing discrimination, and representing disabled people as complex narrative subjects.

References

1. Addlakha, R. (Ed.). (2013). *Disability studies in India: Global discourses, local realities*. Routledge.
2. Ali, M., & Hassan, N. (2022). A survey of computational framing analysis approaches. In Y. Goldberg, Z. Kozareva, & Y. Zhang (Eds.), *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing* (pp. 9335–9348). Association for Computational Linguistics.

3. Dolmage, J. T. (2014). *Disability rhetoric*. Syracuse University Press.
4. Khan, N. Y.-K. (2023). Representation of autism in Vietnamese digital news media: A computational corpus and framing analysis. *Asia Pacific Media Educator, 33*(1), 142–158.
5. Mehrotra, N. (Ed.). (2020). *Disability studies in India: Interdisciplinary perspectives*. Springer.
6. Mitchell, D. T., & Snyder, S. L. (2000). *Narrative prosthesis: Disability and the dependencies of discourse*. University of Michigan Press.
7. Sati, S., Prasad, G. J. V., & Bhattacharjee, R. (Eds.). (2022). *Reclaiming the disabled subject: Representing disability in short fiction* (Vol. 1). Bloomsbury India.
8. Adams, R., Reiss, B., & Serlin, D. (Eds.). (2015). *Keywords for disability studies*. NYU Press.
9. United Nations. (n.d.). *Disability-inclusive language guidelines*. United Nations.
10. Garland-Thomson, R. (1997). *Extraordinary bodies: Figuring physical disability in American culture and literature*. Columbia University Press.
11. Oliver, M. (1990). *The politics of disablement*. Macmillan.
12. Quayson, A. (2007). *Aesthetic nervousness: Disability and the crisis of representation*. Columbia University Press.
13. Shakespeare, T. (2006). *Disability rights and wrongs*. Routledge.
14. Siebers, T. (2008). *Disability theory*. University of Michigan Press.