1

Physiotherapy Patient Records Enhanced with Natural Language Processing

The combination of natural language processing (NLP) and physiotherapy patient records represents a significant step toward tailored care, an improved clinical experience and enhanced evidence-based decision-making.

Physiotherapy teams receive massive volumes of unstructured data, such as progress notes, assessment reports, rehabilitation plans and patient comments, which are difficult to process automatically; as a result, much of this data remains unexploited. NLP is capable of extracting, structuring and processing this text to provide clinicians with valuable information to customize rehabilitation programs.

NLP captures vital clinical data such as range of motion, pain descriptors, functional limitations and patient adherence to therapies, making it easier to monitor patient progress. Integrating NLP with electronic health records (EHRs) improves information exchange, allowing physiotherapists to collaborate more effectively with other healthcare professionals. Additional emerging applications include outcome prediction, real-time clinical decision support and progress tracking through wearable devices connected to documentation systems.

However, privacy concerns, data quality issues and algorithmic bias remain significant challenges. Strict policies and rigorous clinical validation are therefore required. Overall, when properly integrated, physiotherapy and NLP can accelerate research processes and enable precision rehabilitation, where treatment is adjusted in real-time according to the needs of individual patients.

In general, NLP-driven physiotherapy systems can transform narrative notes into actionable information, helping move healthcare toward a more intelligent and data-driven model.

Chapter written by Mandar MALAWADE and Rasika Ranjit CHAFLE.

1.1. Introduction

Physiotherapy is vital in healthcare practice, serving patients with musculoskeletal trauma, neurological conditions, cardiopulmonary complications and post-surgical rehabilitation needs. It aims to restore movement, relieve pain and enhance functionality, thereby improving the quality of life for many patients. Its success depends on comprehensive and consistent recording of patient background, functional status, evaluation, progress and outcomes (Grout et al. 2024).

Physiotherapy data is typically narrative in nature, in the form of progress notes, patient narratives and treatment plan descriptions. Unlike laboratory data or medical images, it is therefore difficult to systematically analyze.

Hospitals and clinics across the globe now use digital health records. These systems are increasingly used in physiotherapy, resulting in an ever-expanding pool of patient data (Siontis et al. 2021). This information contains both structured data, such as demographics, vital signs and diagnosis codes, and unstructured data, such as therapist notes, patient reports and functional goals. The use of physiotherapy records in research, quality improvement and personalized care is often limited because narrative sections are more difficult to process than structured data (Wulff et al. 2020).

Natural language processing (NLP) is a branch of artificial intelligence (AI) that allows computers to interpret human language. It enables machines to extract meaning, organize information and understand context within text, creating opportunities to unlock the value contained in physiotherapy records (Sivarajkumar et al. 2024a). NLP is already widely used in healthcare for tasks such as automated coding, detection of adverse events, analysis of patient sentiment and clinical decision support (Bobba et al. 2023). Extending these capabilities to physiotherapy has the potential to transform rehabilitation practice.

1.1.1. *Why physiotherapy needs NLP*

Physiotherapy records are more complex and distinct compared with many other clinical notes. They include qualitative pain assessments, descriptions of movement restrictions, self-reported fatigue and therapist reports of patient motivation and compliance. These details cannot be adequately represented using numbers alone, yet they are essential for customizing rehabilitation (Yan et al. 2024).

For example, two patients with the same diagnosis may require very different treatment plans depending on psychosocial factors, adherence and progress, which are often documented in narrative notes. NLP has the potential to extract valuable information from this text, allowing clinicians and researchers to identify patterns, predict outcomes and tailor care strategies (Fu et al. 2020).

Physiotherapy also commonly involves collaboration among physicians, occupational therapists, psychologists and nurses. Effective communication requires transforming physiotherapy notes into interoperable structured formats. NLP can automate this structuring process, thereby improving interdisciplinary collaboration and continuity of care (Wu et al. 2020).

1.1.2. *Emerging opportunities*

The addition of NLP to physiotherapy records brings numerous opportunities. First, NLP-based methods, such as entity recognition, text classification and sentiment analysis, can retrieve clinically significant information, including symptoms, goals and rehabilitation milestones. Second, NLP-based decision support may assist therapists in prioritizing interventions, identifying high-risk patients and monitoring adherence over time. Third, when NLP is combined with wearable or motion-capture data, it can provide a more comprehensive picture, as objective measures can be linked with subjective reports, leading to a better understanding of recovery processes (Sivarajkumar et al. 2024b).

1.1.3. *Challenges and need for research*

Despite its potential, NLP in physiotherapy is still in its early stages. Limitations include the absence of physiotherapy-specific annotated corpora, variation in documentation across institutions and concerns related to privacy and patient consent.

Algorithms trained on general medical text may fail to capture the nuances of physiotherapy terminology and therefore require domain adaptation. Close clinical validation is also necessary to ensure that NLP results align with professional expertise and support therapists rather than replace clinical judgment.

1.2. Physiotherapy patient records: structure, content and challenges

Physiotherapy records play a central role in rehabilitation practice. As instruments of clinical communication and medico-legal documentation, they contain details of patient history, assessment, intervention and outcomes. Because they are largely narrative, integrating them with NLP presents both opportunities and challenges, forming the foundation for improved integration with computational approaches such as NLP (Gao et al. 2022).

1.2.1. *Structure of physiotherapy records*

Physiotherapy records generally include three main types of information:

– *Patient demographics and medical history*: these sections document age, gender, current conditions, previous surgeries, medications and baseline functioning. Although these data help establish the clinical background, they do not provide detailed therapeutic information (Zhang et al. 2020).

– *Assessment and evaluation*: these narratives contain patient-reported outcomes (PROs), therapist observations, pain scores and detailed musculoskeletal or neurological assessments. The SOAP format (Subjective, Objective, Assessment, Plan) is widely used, and the subjective and assessment sections typically contain the most textual content (Chua et al. 2023).

– *Treatment plans and progress notes*: therapists record exercise prescriptions, manual therapies, patient compliance and treatment progress. These entries frequently describe variations in pain, movement and psychosocial considerations in free text rather than using formal codes (Tsai et al. 2025).

This combination of structured fields and free text makes physiotherapy records particularly suitable for NLP. Easily extractable variables can be complemented by the richer insights contained in narrative notes through automated analysis.

1.2.2. *Nature of unstructured data*

Unstructured physiotherapy data include patient narratives, therapist observations and discussions of treatment goals. Unlike diagnostic codes, these narratives are contextualized and highly personalized.

For example, a therapist might write: "Patient complains of *difficulty climbing* stairs because her left knee *feels* unstable, *particularly* in the evenings." Such statements capture clinically meaningful information (activity limitations, laterality and symptom variation) that may be overlooked in structured documentation. NLP can extract this information and enable more detailed analysis (Boag et al. 2021).

PROs, such as fatigue, anxiety or adherence issues, are often not recorded in separate fields but embedded within free text. Since recovery is influenced by physical, psychological and social factors, extracting PROs using NLP can enrich patient profiles and improve outcome prediction (Gao et al. 2022).

1.2.3. *Documentation variability*

A persistent challenge is the variation in documentation among clinicians, institutions and regions. Some therapists write detailed narratives, while others use brief checklists or short phrases. Such inconsistency reduces standardization and makes it more difficult to train reliable NLP models (Zhang et al. 2020).

Differences in terminology also create challenges. For example, ROM versus range of motion, or manual therapy versus hands-on treatment, require complex preprocessing strategies, including synonym mapping and contextual embeddings.

Variations among EHR systems across institutions further complicate the issue. Some systems provide structured templates, while others rely heavily on unstructured notes. As a result, physiotherapy corpora remain relatively limited compared with those in fields such as radiology or oncology (Tsai et al. 2025).

1.2.4. *Challenges of current physiotherapy records*

There are several obstacles to the effective use of physiotherapy records in computational analysis:

– *Unstructured data overload*: much of the valuable clinical information exists in free text, making it difficult to query without NLP.

– *Irregular documentation*: incomplete or inconsistent documentation may occur due to time constraints, leading to therapists omitting important details regarding patient progress.

– *Subjectivity of observations*: impressionistic statements (e.g. patient motivated, gait seems improved) in notes are difficult to quantify.

– *Absence of standardized terminology*: there is limited routine use of coding systems such as the International Classification of Functioning (ICF).

– *Poor interoperability of datasets*: physiotherapy notes are often siloed and therefore not easily integrated into multidisciplinary care.

All of these problems contribute to the need for technologies such as NLP, which can transform narrative documentation into structured, analyzable data that can support clinical decision-making (Boag et al. 2021).

1.2.5. *Implications for NLP integration*

The challenges associated with physiotherapy records provide strong arguments in favor of NLP implementation:

– convert free-text progress notes into structured clinical variables;

– identify patterns of symptom development across large patient populations;

– standardize terminology through automated synonym recognition;

– facilitate interoperability by mapping physiotherapy terms to standardized coding systems.

In this way, NLP serves two key objectives: increasing clinical efficiency (reducing documentation burden and improving information retrieval) and enhancing research through large-scale analysis of rehabilitation outcomes.

1.3. Fundamentals of NLP in healthcare

Natural language processing (NLP) is a subdivision of artificial intelligence (AI) that enables machines to read, understand and generate human language. In healthcare, it is primarily used to transform clinical narratives into analyzable data. Because much medical knowledge exists in free-text documents such as discharge summaries, progress notes and pathology reports, NLP has become indispensable to precision medicine. Its importance has increased as electronic health records (EHRs) continue to expand rapidly (Nogales et al. 2024).

1.3.1. *Core components of NLP*

The NLP pipeline typically involves several sequential processes that transform raw text into structured outputs:

– *Text preprocessing*: preprocessing involves tokenization (splitting text into words or phrases), lemmatization or stemming (reducing words to their base forms), and stop-word removal (eliminating common words such as "the" or "and"). In physiotherapy records, preprocessing is complicated by domain-specific abbreviations (e.g. ROM for range of motion, ADLs for activities of daily living) (Noh and Kavuluru 2021).

– *Named entity recognition (NER)*: NER identifies clinically relevant entities such as body parts, symptoms, functional limitations and therapy types. For example, in a note stating "*the* patient *has difficulty walking due to* pain in the right knee", an NLP model may identify walking as a functional activity, right knee as an anatomical location and pain as a symptom.

– *Part-of-speech (POS) tagging and parsing*: this step assigns grammatical roles to words and helps identify relationships between entities. In physiotherapy documentation, this may help link functional activities (e.g. climbing stairs) with descriptions of difficulty, thereby providing a structured representation of functional impairment.

– *Text classification*: text classification algorithms assign notes to predefined categories such as assessment, intervention or outcome. In rehabilitation research, classification can be used to automatically code patient progress into categories such as improved mobility, persistent pain or non-compliance.

– *Sentiment and context analysis*: although sentiment analysis originated in consumer analytics, it is increasingly applied in healthcare. Physiotherapy documentation often includes subjective statements (e.g. patient *highly motivated* or limited engagement). NLP can quantify these insights to help guide adherence strategies (Chen et al. 2020).

– *Information extraction and summarization*: NLP can also generate summaries of long clinical notes. Such systems highlight important events such as symptom onset, therapy modifications and functional milestones, thereby reducing documentation burden and improving clinical efficiency.

1.3.2. *NLP approaches in healthcare*

Two major paradigms dominate healthcare NLP:

– *Rule-based systems*: early NLP systems relied on manually written rules and dictionaries. Experts defined patterns to identify key terms. For example, a rule might detect phrases indicating reduced range of motion. While effective in limited domains, such techniques struggle with ambiguity, synonyms and evolving terminology (Kumar et al. 2026a).

– *Machine learning/deep learning models*: advances in machine learning – particularly deep neural networks – have significantly transformed NLP. Models such as conditional random fields (CRFs), recurrent neural networks (RNNs) and transformer architectures (e.g. BERT, BioBERT and ClinicalBERT) are trained on large corpora to learn semantic patterns. In healthcare applications, domain-specific models trained on medical texts generally outperform generic language models (Agaronnik et al. 2020).

Domain-specific adaptation is particularly important in physiotherapy. General medical datasets may not include rehabilitation-specific expressions such as single-leg stance, asymmetrical gait or functional training. Therefore, accurate analysis requires models that are tailored to physiotherapy data.

1.3.3. *Unique challenges of clinical NLP*

Healthcare NLP faces several challenges that are less common in general NLP:

– *Medical language ambiguity*: clinical notes contain many abbreviations that vary across hospitals (e.g. PT may mean physiotherapy or prothrombin time).

– *Data sparsity*: many physiotherapy concepts are underrepresented in large medical datasets.

– *Domain expertise requirement*: accurate NLP systems require close collaboration with clinicians. Misinterpretation of clinical language may lead to unsafe recommendations.

– *Privacy and ethics*: patient records are confidential and must comply with regulations such as HIPAA or GDPR.

1.3.4. *Importance of NLP in physiotherapy*

Narrative documentation plays a crucial role in physiotherapy. Progress notes frequently include subjective experiences (e.g. stiffness less in the mornings), functional limitations (e.g. *difficulty* climbing stairs) and qualitative observations (e.g. more *stable* walking). These details are essential for therapy decision-making but are difficult to represent in structured systems (Noh and Kavuluru 2021).

NLP helps bridge this gap by:

– converting free-text descriptions into standardized outcomes (e.g. *pain intensity, functional mobility*);

– enabling longitudinal tracking of symptoms and adherence across multiple visits;

– supporting outcome prediction by identifying patterns in narrative clinical data;

– facilitating interoperability by mapping notes to standardized coding systems such as the International Classification of Functioning, Disability and Health (ICF).

For example, an NLP system could process hundreds of physiotherapy notes from stroke patients, extract references to balance, gait and upper-limb function, and track improvements over time. These structured outputs can support patient monitoring and individualized treatment planning (Chen et al. 2020).

1.3.5. *Evolution of NLP tools in clinical practice*

Recent developments include:

– *ClinicalBERT and BioBERT*: transformer-based models pre-trained on medical texts that can interpret clinical language more accurately.

– *cTAKES:* an open-source system that performs entity recognition in EHRs.

– *MedLEE*: an information extraction system designed to convert clinical narratives into structured data.

Although these tools have mainly been applied in fields such as oncology, radiology and cardiology, their application in physiotherapy remains limited. However, the growing availability of rehabilitation data, including PRO measures, creates opportunities for developing specialized NLP models for physiotherapy (Kumar et al. 2026a).

1.3.6. *Future outlook*

The development of NLP in healthcare is moving from rule-based approaches toward context-aware models capable of understanding complex, domain-specific language.

As physiotherapy becomes increasingly digitized, integrating NLP into routine clinical workflows will help unlock the value of unstructured clinical notes. Furthermore, multimodal AI – combining NLP with imaging, sensor data and wearable technologies – offers new opportunities for advanced rehabilitation analytics.

1.4. Applications of NLP in physiotherapy records

The application of NLP to physiotherapy records is still emerging; however, it shows significant potential across several areas of rehabilitation practice.

NLP is particularly well suited for analyzing descriptive and narrative documentation. By extracting actionable insights from free-text data, NLP can support personalized rehabilitation, improve clinical efficiency and enhance patient engagement (Kumar et al. 2026b).

The following sections explore key applications of NLP in physiotherapy. Table 1.1 presents an overview of these applications.

Application area	Description	Examples from paper
Information extraction and structuring	Converts free-text physiotherapy notes into structured variables by identifying symptoms, functional limitations, goals and progress markers	Extracting "hip flexion", "difficulty with standing", mapping notes to ICF classifications
Clinical decision support	NLP tools help therapists make informed decisions by highlighting trends, risks and intervention opportunities	Alerting increased pain, recommending typical regimens for TKR patients, triaging high-risk individuals
Predictive analytics & outcome monitoring	Identifies patterns in narrative notes to forecast recovery, detect risks and compare intervention effectiveness	Predicting ADL independence, flagging poor adherence or fatigue mentions
Integration with wearables/sensor data	Merges subjective notes with objective movement and sensor data for comprehensive patient monitoring	Combining step count with notes such as "difficulty with stair climbing"
Patient engagement and communication	Uses NLP to analyze motivation, personalize messages and interact via chatbots	Sentiment analysis, virtual assistants reminding exercises
Research and population-level insights	Enables large-scale mining of physiotherapy datasets to identify trends, disparities and outcomes	Identifying adherence barriers in older adults, women or rural populations

Table 1.1. *Applications of NLP in physiotherapy records*

1.4.1. *Information extraction and structuring*

One of the most fundamental applications of NLP is the extraction of clinically relevant information from unstructured records. Physiotherapy notes often describe symptoms, functional limitations, pain trajectories and therapy responses in free-text format.

For instance, a typical note might read: "*Patient demonstrates improved hip flexion but continues to report difficulty with prolonged standing.*" In such cases, NLP can automatically identify elements such as hip flexion (movement parameter), difficulty standing (functional limitation) and improved (progress indicator).

Through information extraction, it is possible to achieve:

– *Standardization of clinical notes*: free-text documentation can be mapped to controlled vocabularies such as the International Classification of Functioning (ICF).

– *Longitudinal tracking*: repeated mentions of improvement or decline across sessions can be aggregated to model patient recovery trajectories.

– *Automated summaries*: NLP can generate concise progress reports for multidisciplinary teams, reducing the burden of manual summarization (Al Khatib et al. 2024).

Importantly, structured outputs allow physiotherapists to query patient records more efficiently. In both clinical practice and research, therapists can search through hundreds of notes using keywords such as knee instability during gait and quickly identify relevant patient subsets.

1.4.2. *Clinical decision support*

Clinical decision-making in physiotherapy integrates functional assessments, subjective reports and therapeutic goals. NLP can structure this information to support evidence-based recommendations (Poongodi et al. 2020).

Use cases in decision support

– *Treatment recommendation*: NLP models trained on large physiotherapy datasets can identify common patterns of treatment choices, such as exercise programs used for patients after total knee replacement.

– *Alert systems*: NLP can detect warning signals, such as persistent or increasing pain despite treatment, prompting early intervention.

– *Coordination with multidisciplinary teams*: standardized terminology improves communication among physiotherapists, physicians and occupational therapists.

Decision-support systems become particularly valuable when therapists manage high caseloads. AI-assisted triage and prioritization can help reduce workload and improve care efficiency.

1.4.3. *Predictive analytics and outcome monitoring*

Beyond information extraction, NLP also supports predictive analytics. By analyzing longitudinal clinical notes, NLP can identify patterns that help predict rehabilitation outcomes and potential complications (Gumiel et al. 2021).

Examples of predictive applications

– *Recovery prediction*: NLP can analyze textual descriptions of functional milestones to estimate the likelihood that a patient will regain independence in activities of daily living (ADLs).

– *Risk stratification*: notes that repeatedly mention fatigue, poor adherence or psychosocial barriers can be flagged as indicators of potentially poor rehabilitation outcomes.

– *Comparative effectiveness research*: NLP enables large-scale analysis of rehabilitation outcomes across different therapy approaches by mining text data for associations between interventions and outcomes.

For example, in stroke rehabilitation, NLP can track how frequently terms such as upper-limb function or balance training appear in clinical notes and correlate them with recovery outcomes to help identify the most effective interventions.

1.4.4. *Incorporation with wearables and sensor data*

Patient progress is increasingly monitored using wearable sensors, motion sensor and mobile applications. These technologies provide quantitative parameters such as step count and gait asymmetry but offer limited insight into patient experiences. NLP helps bridge this gap by analyzing clinical notes that describe pain, fatigue or exercise adherence (Alexander et al. 2022).

Multimodal integration

– *Integration of NLP with wearable metrics*: combining wearable sensor data with NLP-derived insights provides a more comprehensive progress profile. For example, a wearable device may record increased step counts, while clinical notes indicate that the patient still experiences difficulty climbing stairs.

– *Real-time feedback*: NLP analysis of patient diaries or app-based messages can provide clinicians with continuous updates on symptoms and barriers to recovery.

– *Remote rehabilitation*: in tele-physiotherapy settings, NLP can analyze text or voice messages from patients to complement sensor data, enabling remote monitoring of adherence and treatment outcomes.

This multimodal approach allows for the development of more accurate, patient-centered recovery models by integrating both objective and subjective data.

1.4.5. *Patient engagement and communication*

Successful rehabilitation depends not only on the expertise of therapists but also on patient motivation and adherence to treatment plans. NLP can enhance communication and patient engagement:

– *Sentiment analysis of patient notes*: physiotherapy documentation often contains comments regarding patient motivation or adherence (e.g. “highly engaged” or “missed sessions”). NLP can quantify these observations to identify patients who may be at risk of disengagement.

– *Chatbots and virtual assistants*: NLP-driven conversational agents can provide patients with personalized reminders, exercise guidance and motivational support between clinic visits.

– *PROs*: free-text responses from patients – submitted through surveys, mobile applications or diaries – can be analyzed using NLP to monitor patient experiences and satisfaction with therapy.

These applications strengthen the therapist–patient relationship by facilitating shared decision-making and supporting individualized care.

1.4.6. *Research and population-level insights*

NLP enables large-scale analysis of physiotherapy datasets for research purposes. Population-level studies can reveal trends in rehabilitation outcomes, variations in care delivery and patterns in therapy adherence. For example, analyzing thousands of clinical notes may reveal barriers to rehabilitation among specific demographic groups, such as older adults, women or patients in rural areas (Kumar et al. 2026b).

These insights can inform healthcare policy, support quality improvement initiatives and contribute to the education and training of future physiotherapists. By converting narrative clinical documentation into structured data, NLP helps unlock previously underutilized clinical knowledge.

1.5. Technological frameworks for NLP in physiotherapy

The application of NLP in physiotherapy requires robust technological frameworks capable of managing the complexity of rehabilitation documentation while ensuring scalability, security and clinical reliability. Compared with many other clinical NLP applications, physiotherapy presents unique challenges because its documentation is largely narrative, contains numerous abbreviations and frequently focuses on subjective measures such as fatigue, pain, motivation and functional limitations.

Therefore, technological systems must combine advanced NLP models with healthcare information systems while remaining flexible enough to adapt to the language and workflows of physiotherapy practice (Zini and Awad 2022).

1.5.1. *Data acquisition and preprocessing*

The foundation of any NLP system is data acquisition and preprocessing. In physiotherapy, relevant datasets include EHRs, therapist progress notes, patient diaries and PROs.

Several important preprocessing steps are required:

– *Text cleaning*: removing irrelevant information such as headers, timestamps or template-generated text.

– *Normalization*: expanding abbreviations (e.g. ROM → range of motion), resolving synonyms and ensuring terminology consistency.

– *De-identification*: removing protected health information (PHI) to ensure compliance with privacy regulations.

– *Annotation*: creating gold-standard datasets in which expert physiotherapists label clinical concepts, providing training data for supervised learning models.

Physiotherapy-specific corpora remain limited, making annotation particularly important. Semi-supervised learning and transfer learning approaches – where models are first trained on broader clinical corpora and then fine-tuned using smaller physiotherapy datasets – offer a practical solution (Ravi et al. 2025).

1.5.2. *NLP model architectures*

Several NLP systems can be applied in physiotherapy, each with its advantages and limitations:

– *Rule-based systems*: effective for recognizing specific physiotherapy terms but limited in flexibility when handling variations in documentation styles.

– *Classical machine learning models*: algorithms such as support vector machines (SVMs) and random forests can classify physiotherapy notes into categories (e.g. progress, intervention, adherence).

– *Deep learning architectures*: neural networks and transformer-based architectures dominate contemporary clinical NLP. Models such as BERT, BioBERT, ClinicalBERT and potential adaptations such as RehabBERT can learn semantic nuances in rehabilitation language.

– *Hybrid systems*: combining rule-based lexicons (for physiotherapy-specific terminology) with deep learning models can improve accuracy, particularly in specialized rehabilitation domains (Zini and Awad 2022).

1.5.3. *EHR integration*

To achieve clinical impact, NLP systems must be effectively integrated into existing electronic health record (EHR) systems. Key considerations include:

– *Interoperability*: standardizing information according to frameworks such as HL7 FHIR (Fast Healthcare Interoperability Resources) ensures that NLP outputs can be shared across healthcare systems.

– *On-demand processing*: by embedding NLP modules within EHR systems, therapist notes can be structured automatically at the time of documentation, reducing administrative workload.

– *User interface design*: visual dashboards can present NLP-derived insights – such as functional progress trends, adherence risks or outcome predictions – in therapist-friendly formats.

For example, an EHR-integrated NLP module may detect repeated mentions of gait instability in clinical notes and flag the patient as being at risk of falls, prompting preventive interventions (Mansouri 2025).

1.5.4. *Cloud-based and federated learning approaches*

Privacy regulations related to medical data often restrict the sharing of physiotherapy records across institutions, limiting the development of large-scale NLP models. Cloud-based systems and federated learning approaches help address this challenge:

– *Cloud-based NLP*: modular systems allow hospitals to send de-identified clinical notes to secure cloud platforms, where they are processed using powerful NLP models.

– *Federated learning*: models are trained locally within institutions, and only model parameters – not raw patient data – are shared. This approach enables the development of robust, generalizable NLP systems without compromising patient privacy.

These approaches are particularly valuable in physiotherapy, where data are often distributed across small clinics and rehabilitation centers.

1.5.5. *Multimodal integration frameworks*

Recent developments in AI emphasize multimodal integration, which combines NLP with other sources of data. In physiotherapy, this includes:

– *Wearables and IoT devices*: motion sensors, smartwatches and pressure-sensitive insoles provide objective movement data. NLP adds contextual information from therapist notes and patient diaries.

– *Imaging data*: NLP can extract textual descriptions (e.g. therapist observations of gait deviations) and correlate them with video-based gait analysis or posture-recognition systems.

– *Patient feedback platforms*: NLP can analyze free-text responses from surveys or mobile applications and combine them with sensor data to create a comprehensive view of patient recovery.

This integration of structured and unstructured data supports precision rehabilitation, where therapy plans can be dynamically adjusted based on combined insights from multiple data sources (Ravi et al. 2025).

1.5.6. *Governance, security and ethical considerations*

Technological frameworks must also address governance, security and ethical considerations. Key requirements include:

– *Data protection*: ensuring compliance with regulations such as HIPAA, GDPR and local health data protection laws.

– *Bias reduction*: NLP models should be trained using diverse physiotherapy datasets to avoid reinforcing inequities in care.

– *Explainability*: therapists must understand why an NLP system generates a particular recommendation; therefore, explainable AI methods are essential.

– *Sustainability*: frameworks should be scalable, cost-efficient and adaptable to resource-constrained healthcare environments.

Without strong governance mechanisms, clinician trust and patient safety may be compromised (Mansouri 2025).

1.6. Clinical benefits and opportunities

The integration of NLP into physiotherapy records has the potential to transform physiotherapy practice. Beyond the technical capabilities of NLP models, their greatest value lies in improving patient outcomes, streamlining clinical workflows and enabling data-driven decision-making. Although adoption is still in its early stages, several clinical benefits present significant opportunities for physiotherapists, patients and healthcare systems (Davids et al. 2022).

1.6.1. *Enhancing clinical decision-making*

One of the most important advantages of NLP integration is improved clinical decision-making. Physiotherapists often face the challenge of synthesizing large volumes of narrative information recorded across multiple patient visits. By organizing unstructured clinical notes, NLP enables therapists to:

– *Identify patient-specific trends*: repeated mentions of functional limitations, such as difficulty climbing stairs or persistent pain, can be automatically highlighted, supporting more targeted interventions.

– *Enable evidence-based practice*: NLP can connect textual descriptions of interventions (e.g. manual therapy or neuromuscular re-education) with reported outcomes, enabling evidence-based comparisons.

– *Support multidisciplinary collaboration*: structured outputs make physiotherapy data interoperable with physician, nursing and occupational therapy documentation, supporting more comprehensive care planning.

For example, a physiotherapist working with a patient recovering from hip replacement surgery may use NLP-generated summaries that highlight improvements in mobility alongside persistent pain symptoms, enabling more precise adjustments to the therapy plan (Rajpurkar et al. 2022).

1.6.2. *Reducing documentation burden*

Documentation is frequently cited as one of the most time-consuming aspects of physiotherapy practice. Time spent on manual charting can reduce the time available for direct patient care and contribute to professional burnout.

NLP tools can reduce this burden by:

– *Automating note summarization*: long clinical narratives can be condensed into concise summaries that capture key milestones without unnecessary detail.

– *Voice-to-text with contextual structuring*: NLP-powered dictation systems allow therapists to record observations verbally, which are then converted into structured clinical documentation automatically (Agaronnik et al. 2020).

– *Standardized templates*: by automatically inserting commonly used terminology and formatting notes according to regulatory standards, NLP reduces repetitive documentation tasks.

Overall, NLP tools can reduce administrative workload by automatically organizing relevant clinical information, thereby saving time while improving the completeness and consistency of patient records – both of which are essential for high-quality care and medico-legal compliance.

1.6.3. *Improving patient engagement*

Patient engagement is a critical factor in successful rehabilitation. Adherence to prescribed exercises and attendance at therapy sessions are strongly associated with treatment outcomes. NLP technologies can enhance patient communication through conversational systems and personalized feedback mechanisms:

– *Personalized feedback systems*: NLP can analyze patient self-reports submitted through mobile applications and generate personalized feedback messages that encourage adherence.

– *Chatbots and virtual assistants*: NLP-powered conversational agents can answer patient questions, remind them of exercise schedules and monitor adherence remotely.

– *Tracking motivation*: sentiment analysis of therapist notes can quantify patient motivation, helping clinicians identify individuals who may be at risk of disengagement (Himani and Pathak 2024).

These tools support a more patient-centered approach, extending communication beyond the clinic into the patient's everyday life.

1.7. Conclusion

The integration of NLP with physiotherapy patient records represents an important advancement in rehabilitation practice. Physiotherapy is a documentation-intensive discipline that relies heavily on narrative descriptions of impairments, progress and therapeutic goals. NLP transforms these unstructured texts into structured, actionable information, enabling physiotherapists to retrieve information more efficiently, make informed clinical decisions and deliver more personalized care.

At the clinical level, NLP supports decision-making by identifying patterns of patient improvement and linking interventions with outcomes. It also reduces documentation burden through automated summaries and voice-to-text technologies, allowing clinicians to devote more time to direct patient care. In addition, NLP enhances patient engagement by enabling conversational agents, personalized feedback systems and tools for monitoring motivation and adherence.

Beyond individual patient care, NLP also contributes to predictive and preventive healthcare by identifying risk factors early, supporting fall-risk stratification and helping prevent complications. At the population level, it transforms physiotherapy records into valuable resources for research, quality improvement and health policy planning. These developments support the emergence of precision rehabilitation, where interventions are dynamically adapted to the individual needs and recovery profiles of patients.

1.8. References

Agaronnik, N.D., Lindvall, C., El-Jawahri, A., He, W., Iezzoni, L.I. (2020). Challenges of developing a natural language processing method with electronic health records to identify persons with chronic mobility disability. *Archives of Physical Medicine and Rehabilitation*, 101(10), 1739–1746.

Al Khatib, H.S., Neupane, S., Kumar Manchukonda, H., Golilarz, N.A., Mittal, S., Amirlatifi, A., Rahimi, S. (2024). Patient-centric knowledge graphs: A survey of current methods, challenges, and applications. *Frontiers in Artificial Intelligence*, 7, 1388479.

Alexander, G., Bahja, M., Butt, G.F. (2022). Automating large-scale healthcare service feedback analysis: Sentiment analysis and topic modeling study. *JMIR Medical Informatics*, 10(4), e29385.

Boag, W., Kovaleva, O., McCoy Jr, T.H., Rumshisky, A., Szolovits, P., Perlis, R.H. (2021). Hard for humans, hard for machines: Predicting readmission after psychiatric hospitalization using narrative notes. *Translational Psychiatry*, 11(1), 32.

Bobba, P.S., Sailer, A., Pruneski, J.A., Beck, S., Mozayan, A., Mozayan, S., Arango, J., Cohan, A., Chheang, S. (2023). Natural language processing in radiology: Clinical applications and future directions. *Clinical Imaging*, 97, 55–61.

Campagnini, S., Arienti, C., Patrini, M., Liuzzi, P., Mannini, A., Carrozza, M.C. (2022). Machine learning methods for functional recovery prediction and prognosis in post-stroke rehabilitation: A systematic review. *Journal of NeuroEngineering and Rehabilitation*, 19(1), 54.

Chen, X., Xie, H., Cheng, G., Poon, L.K., Leng, M., Wang, F.L. (2020). Trends and features of the applications of natural language processing techniques for clinical trials text analysis. *Applied Sciences*, 10(6), 2157.

Chua, M., Kim, D., Choi, J., Lee, N.G., Deshpande, V., Schwab, J., Lev, M.H., Gonzalez, R.G., Gee, M.S., Do, S. (2023). Tackling prediction uncertainty in machine learning for healthcare. *Nature Biomedical Engineering*, 7(6), 711–718.

Davids, J., Lidströmer, N., Ashrafian, H. (2022). Artificial intelligence for physiotherapy and rehabilitation. In *Artificial Intelligence in Medicine*, Lidströmer, N. and Ashrafian, H. (eds). Springer, Cham, 1789–1807.

Fu, S., Chen, D., He, H., Liu, S., Moon, S., Peterson, K.J., Shen, F., Wang, L., Wang, Y., Wen, A. et al. (2020). Clinical concept extraction: A methodology review. *Journal of Biomedical Informatics*, 109, 103526.

Gao, C., Goswami, M., Chen, J., Dubrawski, A. (2022). Classifying unstructured clinical notes via automatic weak supervision. In *Machine Learning for Healthcare Conference*. PMLR, 673–690.

Grout, R., Gupta, R., Bryant, R., Elmahgoub, M.A., Li, Y., Irfanullah, K., Patel, R.F., Fawkes, J., Inness, C. (2024). Predicting disease onset from electronic health records for population health management: A scalable and explainable Deep Learning approach. *Frontiers in Artificial Intelligence*, 6, 1287541.

Gumiel, Y.B., Silva e Oliveira, L.E., Claveau, V., Grabar, N., Paraiso, E.C., Moro, C., Carvalho, D.R. (2021). Temporal relation extraction in clinical texts: A systematic review. *ACM Computing Surveys (CSUR)*, 54(7), 1–36.

Himani, P.K.V., Pathak, P. (2024). Personalized predictive modeling for rehabilitation outcomes in neurological conditions through machine learning. *Frontiers in Health Informatics* 13(3), 9227–9238.

Kumar, A., Batta, P., Ahuja, S., Rathore, P.S. (eds) (2026a). *AI-driven Innovations in Physiotherapy and Oncology 1*. ISTE Ltd, London and Wiley, Hoboken, NJ.

Kumar, A., Batta, P., Ahuja, S., Rathore, P.S. (eds) (2026b). *AI-driven Innovations in Physiotherapy and Oncology 2*. ISTE Ltd, London and Wiley, Hoboken, NJ.

Mansouri, B. (2025). Survey of abstract meaning representation: Then, now, future. *arXiv preprint*.

Nogales, A., Rodríguez-Aragón, M., García-Tejedor, Á.J. (2024). A systematic review of the application of deep learning techniques in the physiotherapeutic therapy of musculoskeletal pathologies. *Computers in Biology and Medicine*, 172, 108082.

Noh, J. and Kavuluru, R. (2021). Improved biomedical word embeddings in the transformer era. *Journal of Biomedical Informatics*, 120, 103867.

Poongodi, T., Sumathi, D., Suresh, P., Balusamy, B. (2020). Deep learning techniques for electronic health record (EHR) analysis. In *Bio-inspired Neurocomputing*, Bhoi, A.K., Mallick, P.K., Liu, C.-M., Balas, V.E. (eds). Springer, Singapore, 73–103.

Rajpurkar, P., Chen, E., Banerjee, O., Topol, E.J. (2022). AI in health and medicine. *Nature Medicine*, 28(1), 31–38.

Ravi, S., C, M., P,H.L., Kumar, A., Singh Rathore, P. (2025). Multiscale feature fusion for automated classification of digital breast tomosynthesis volumes using 3D deep learning. *Discover Computing*, 28(1), 300. doi: 10.1007/s10791-025-09809-6.

Siontis, K.C., Noseworthy, P.A., Attia, Z.I., Friedman, P.A. (2021). Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. *Nature Reviews Cardiology*, 18(7), 465–478.

Sivarajkumar, S., Mohammad, H.A., Oniani, D., Roberts, K., Hersh, W., Liu, H., He, D., Visweswaran, S., Wang, Y. (2024a). Clinical information retrieval: A literature review. *Journal of Healthcare Informatics Research*, 8(2), 313–352.

Sivarajkumar, S., Gao, F., Denny, P., Aldhahwani, B., Visweswaran, S., Bove, A., Wang, Y. (2024b). Mining clinical notes for physical rehabilitation exercise information: Natural language processing algorithm development and validation study. *JMIR Medical Informatics*, 12(1), e52289.

Tsai, M.L., Chen, K.F., Chen, P.C. (2025). Harnessing electronic health records and artificial intelligence for enhanced cardiovascular risk prediction: A comprehensive review. *Journal of the American Heart Association*, 14(6), e036946.

Wu, S., Roberts, K., Datta, S., Du, J., Ji, Z., Si, Y., Soni, S., Wang, Q., Wei, Q., Xiang, Y. et al. (2020). Deep learning in clinical natural language processing: A methodical review. *Journal of the American Medical Informatics Association*, 27(3), 457–470.

Wulff, A., Mast, M., Hassler, M., Montag, S., Marschollek, M., Jack, T. (2020). Designing an openEHR-based pipeline for extracting and standardizing unstructured clinical data using natural language processing. *Methods of Information in Medicine*, 59(S 02), e64–e78.

Yan, C., Ong, H.H., Grabowska, M.E., Krantz, M.S., Su, W.C., Dickson, A.L., Peterson, J.F., Feng, Q., Roden, D.M., Stein, C.M. et al. (2024). Large language models facilitate the generation of electronic health record phenotyping algorithms. *Journal of the American Medical Informatics Association*, 31(9), 1994–2001.

Zhang, D., Yin, C., Zeng, J., Yuan, X., Zhang, P. (2020). Combining structured and unstructured data for predictive models: A deep learning approach. *BMC Medical Informatics and Decision Making*, 20(1), 280.

Zini, J.E., Awad, M. (2022). On the explainability of natural language processing deep models. *ACM Computing Surveys*, 55(5), 1–31.