Outlook

Scientific Reports: Decision on "Proactive Concussion Prediction Using Symmetry-Aware Multiscale Explainable Hybrid Deep Learning and Multimodal Data Fusion"

From Scientific Reports <do-not-reply@springernature.com>
Date Wed 5/20/2026 5:40 PM
To Akeem Kadri <akeem.kadri@kwasu.edu.ng>

Dear Dr Kadri,

Re: "Proactive Concussion Prediction Using Symmetry-Aware Multiscale Explainable Hybrid Deep Learning and Multimodal Data Fusion"

We are delighted to let you know that the above submission, which you co-authored, has been accepted for publication in Scientific Reports.

Please contact the corresponding author if you would like further details on this decision, including any reviewer feedback.

Thank you for choosing Scientific Reports and we look forward to publishing your article.

Kind regards,

Peer Review Advisors
Scientific Reports

Proactive Concussion Prediction Using Symmetry-Aware Multiscale Explainable Hybrid Deep Learning and Multimodal Data Fusion

Akinbowale Nathaniel BABATUNDE[1, [0000-0001-5168-6071]], Damilare Peter OYINLOYE[2], Roseline Oluwaseun OGUNDOKUN[3, *], Folasade Abimbola ALUKO[4], Ayodele BABATUNDE[5], Akeem Femi KADRI[6], Shuaib Babatunde MOHAMMED[7], Damilola POPOOLA[8], Joseph Bamidele AWOTUNDE[9], Rotimi-Williams BELLO[10]

[1]Department of Computer Science, Kwara State University, Malete, Nigeria. akinbowale.babatunde@kwasu.edu.ng
[2]Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway. peter.d.oyinloye@ntnu.no
[3]Department of Computer Systems Engineering, Tshwane University of Technology, Pretoria, South Africa.
[3]Department of Computer Science, Redeemer's University Ede, Osun State, Nigeria.
[3]Department of Software Engineering, Kaunas University of Technology, Kaunas, Lithuania.
[3]Department of Mathematics and Computing, Thomas Adewumi, Oko, Nigeria.
Ogundokunroseline1@gmail.com, ogundokunr@run.edu.ng
[4]Information and Engineering Department, Fuzhou Melbourne Polytechnic, Fuzhou, China. folasadeabimbola@melbournepolytechnic.edu.au
[5]Department of Statistics, Faculty of Physical Sciences, University of Ilorin, Ilorin, 240003, Nigeria linkayodele@gmail.com
[6]Department of Computer Science, Kwara State University, Malete, Nigeria. akeem.kadri@kwasu.edu.ng
[7]Department of Computer Science, Kwara State University, Malete, Nigeria. shuaib.mohammed@kwasu.edu.ng
[8]Department of Computer Science, Kwara State University, Malete, Nigeria. damilola.popoola13@kwasu.edu.ng
[9]Department of Computer Engineering, Faculty of Engineering and Architecture, Recep Tayyip Erdoğan Üniversitesi, Zihni Derin Yerleşkesi, Fener, 53100, Merkez/Rize, Türkiye. josephbamidele.awotunde@erdogan.edu.tr
[10]Department of Computer Systems Engineering, Tshwane University of Technology, Pretoria, South Africa.

[*]Corresponding author: Roseline Oluwaseun Ogundokun, Tshwane University of Technology

Abstract:
Diagnosis of concussion has been hampered by its subtle and late symptoms, the reliance on subjective judgment, and the lack of real-time multimodal data integration. The developed computational models are generally black boxes and lack the necessary robustness to be useful in safety-critical scenarios. We introduce a symmetry-aware multiscale explainable hybrid deep learning framework for early prediction of concussion risk. The system effectively combines kinematic sensor data, EEG, contextual information, and synchronized video using separate modality-specific feature extractors and a multimodal fusion layer that utilizes attention-based multimodality fusion. The framework is evaluated using subject-independent 5-fold cross-validation to evaluate its generalizability. The proposed framework incorporates explainability with SHAP value analysis and attention visualizations. The proposed system achieves accuracy of 0.91, macro F1-score of 0.90, and AUC of 0.95, outperforming baselines based on single modalities and traditional multimodality integration. High concussion risk events are recalled with F1-score of 0.89 and moderate concussion events are predicted with F1-score of 0.87 to capture subtle injury characteristics. We showed that symmetry-aware multiscale fusion, combined with explainable decision support system, provides a useful tool for concussion risk prediction. This framework holds a promise in terms of its potential for clinical decision support, although further clinical trials using larger, more diverse populations are necessary to establish its reliability for real-world implementation.

Keyword: Concussion Prediction, Explainable AI, Transformer Network, Multimodal Fusion, Symmetry – Aware Learning, EEG Analysis

1. Introduction

Concussion has been an ongoing problem in public health, especially within the context of physically demanding and competitive sport activities, as well as strenuous operational conditions where high impact forces can repeatedly be applied to the head. In fact, concussion and minor TBI may result in ambiguous and variable symptoms that can interfere with an immediate diagnosis and increase the probability of long-term damage if such injuries are not addressed early (Dell et al., 2021; Edelstein et al., 2024). Rapid development in AI and intelligent sensing, has opens up a promising avenue for anticipatory concussion risk assessment on top of reactive post-impact clinical testing (Isiaka et al., 2024; Nayak et al., 2025). In terms of application, this study would be related to sports medicine, neuro-trauma monitoring and intelligent healthcare systems, where an early identification of concussion helps mitigate potential long-term neuro-psychological consequences. Advanced helmets with integrated sensing capabilities are capable of continuously logging kinematics of the head, forces being applied, and dynamic behavior during sport or intense physical work (Zhuang et al., 2021; Du et al., 2024), and flexible morphologically adaptive sensing can further enable reliable detection of head impacts under complex, dynamic operation environments (Zompanti et al., 2024). The presence of sensor-equipped cloud systems enables distributed, near real-time acquisition, and processing of biomechanical and physiological signals (Menghani et al., 2023).

Despite these advancements, a number of factors remain that impede the accuracy of current concussion prediction algorithms. Using only biomechanical impact thresholds does not sufficiently capture the nature of neurophysiological responses to injury and thus are not sensitive to concussion of moderate risk. In fact, studies utilizing neuroimaging and neural signals show evidence of structural and functional brain abnormalities, even in the absence of immediately identifiable signs of injury, underscoring the value of objective neurophysiological biomarkers for diagnosis (De Luca et al., 2023). Alone, however, neurophysiological models are also prone to signal noise, context-independence and loss of robustness outside of the laboratory.

Therefore, multimodal learning has become one of the most promising and key techniques in enhancing the reliability of prediction by the fused physiological, biomechanical, and contextual information. Several previous research conducted in the biomedical and clinical prediction have clearly verified that heterogeneity multi-modalities can enhance the accuracy of classification and model generalization than any single-modality based model (Abrol et al., 2019; Liu et al., 2023). In the application of deep learning, attention-based and hybrid deep learning models allow us to achieve both better adaptability and robustness by adaptively focusing on informative modality under various data-quality levels. Meanwhile, some application works using convolution neural networks to perform the real-time monitoring of helmets and safety during athletic activities also showed the potential to develop an intelligent prediction system working in the dynamic operational context (Iparraguirre-Villanueva et al., 2024). Motivated by these developments, this study proposes an attention-driven multimodal deep learning framework that integrates kinematic sensor data, neurophysiological signals, and contextual information for concussion risk prediction. Adaptive fusion mechanisms and explainability-oriented learning strategies are incorporated to improve predictive robustness, interpretability, and reliable concussion-risk assessment within real-world sports and operational environments.

2. Review of Related Works

Existing concussion prediction research increasingly recognizes mild traumatic brain injury as a multidimensional problem requiring integration of biomechanical, neurophysiological, and contextual information for reliable assessment. While traditional clinical measures have some use for neuro screening, they are often not sufficiently sensitive to either subtle or delayed concussive signs especially when re-occurring events take place (Dell et al., 2021; Mahajan et al., 2023; Edelstein et al., 2024). Such shortfalls have resulted in increased interest in the implementation of artificial intelligence and intelligent sensor technology to allow for predictive concussion risk evaluation. Research has recently shown that sensors capable of recording head kinematics, impacts, and force distributions, implemented as wearable sensing systems in high-risk sporting or operating situations, are a step in this direction (Zhuang et al., 2021; AlNuaimi et al., 2024). The development of flexible and embedded sensor technology allows for increased ecological validity and prolonged, constant monitoring in the real world (Zompanti et al., 2024). The direct evidence for concussive changes in brain network activity provided by measures such as EEG and neural imaging offer a complementary view of these events (Koochaki et al., 2020; Salsabilian & Najafizadeh, 2021; Salsabilian et al., 2022). Similarly, clinical research also suggests correlation between tissue strain and subsequent neurological abnormalities (Zhan et al., 2022; Jolly et al., 2021).

In the machine learning and deep learning approaches, unimodal systems have transformed into multimodal fusion architectures that incorporate heterogeneous physiological and contextual data. CNN-, RNN-, LSTM-, and Transformer-based models are explored extensively in the concussion prediction and other related biomedical classification problems. CNN is beneficial for spatial features learning in sensor or imaging data; while recurrent and LSTM based models facilitate temporal sequence learning in physiological signals. Moreover, transformer and attention-based architectures contribute to effective multimodal learning by adaptively weighing informative modalities under noise or incompletion. (Liu et al. 2023; Ye et al. 2023) Many multimodal fusion works reported the superior performance in prediction to single modality. (Asani et al. 2023; Mazumder et al. 2023). Other relevant smart health systems, behavior analytics, oncology and biomedical imaging research also proves the extensibility of multimodal deep learning in various medical applications. (Sprint et al. 2022; Mahootiha et al. 2025; Nayak et al. 2025; Vylala et al. 2025)

Despite these advances, several limitations remain evident in current literature. Most existing multimodal systems heavily rely on delayed fusion schemes, therefore, compromise real-time applicability for live sporting and operational settings. Most existing deep learning methods are black boxes, which lower physician confidence and practical reliability. Also, cross population, sensor configuration and operational condition generalization are still problematic due to restricted training datasets and assumptions on domain. Similarly, low computational feasibility and robustness to missing modalities, along with few evaluations under real conditions, hinder real-world applications (Isiaka et al., 2024; Tran et al., 2024). Representative studies and relevant contributions/limitations were provided in Table 1. Comparative analysis in Table 1 highlights persistent gaps involving interpretability, real-time adaptability, multimodal robustness, and concussion-specific validation.

Table 1: Summary of Related Theoretical Models

Author (Year)	Model	Algorithm	Contribution	Gaps
Abrol et al. (2019)	Multimodal Fusion Model	Deep Neural Networks	Effective fusion of dynamic connectivity features	Limited interpretability
Asani et al. (2023)	BioscoreNet	Multimodal Self-Attention	Sensor-based TBI detection using attention fusion	Smartphone-limited sensing
Liu et al. (2023)	Attention-based Fusion	Multimodal Attention	Robust prediction under missing modalities	Not concussion-specific
Nishizawa et al. (2025)	Attention Fusion Model	Multimodal Deep Learning	Improved interpretability in clinical prediction	Limited neurotrauma validation
Tran et al. (2024)	Uncertainty-aware Model	Bayesian Deep Learning	Predictive uncertainty modeling	Limited multimodal integration
Ghonge et al. (2025)	Cycle-aware Model	Hybrid Deep Learning	Physiological variability adaptation	Not applied to concussion
Babatunde et al. (2025a)	CNN-based Classifier	DenseNet	Improved medical image classification	No multimodal fusion
Babatunde et al. (2025b)	Transfer Learning Model	MobileNet	Enhanced computational efficiency	Single-domain focus
Mazumder et al. (2023)	Multimodal DL Model	Hybrid Deep Learning	Pediatric mTBI detection using multimodal data	Limited real-time deployment
Mahootiha et al. (2025)	Multimodal Risk Model	Deep Neural Networks	Improved recurrence risk prediction	Oncology-specific application
Ye et al. (2023)	Hybrid Fusion Framework	Multimodal NLP + DL	Enhanced clinical prediction models	Limited sensor integration

From the literature reviewed, we highlight three key research gaps:

1. Limited real-time multimodal integration under dynamic operational settings.
2. Poor interpretability and explainability of current deep learning methods.
3. Limited generalization over populations, sensing devices and deployment conditions.

Figure 1 illustrates the linkage between the above research gaps and the proposed framework components; attention-based multimodal fusion, explainable AI techniques, and generalization-aware validation mechanisms work together to overcome such challenges in real-time performance, robustness, and deployability.

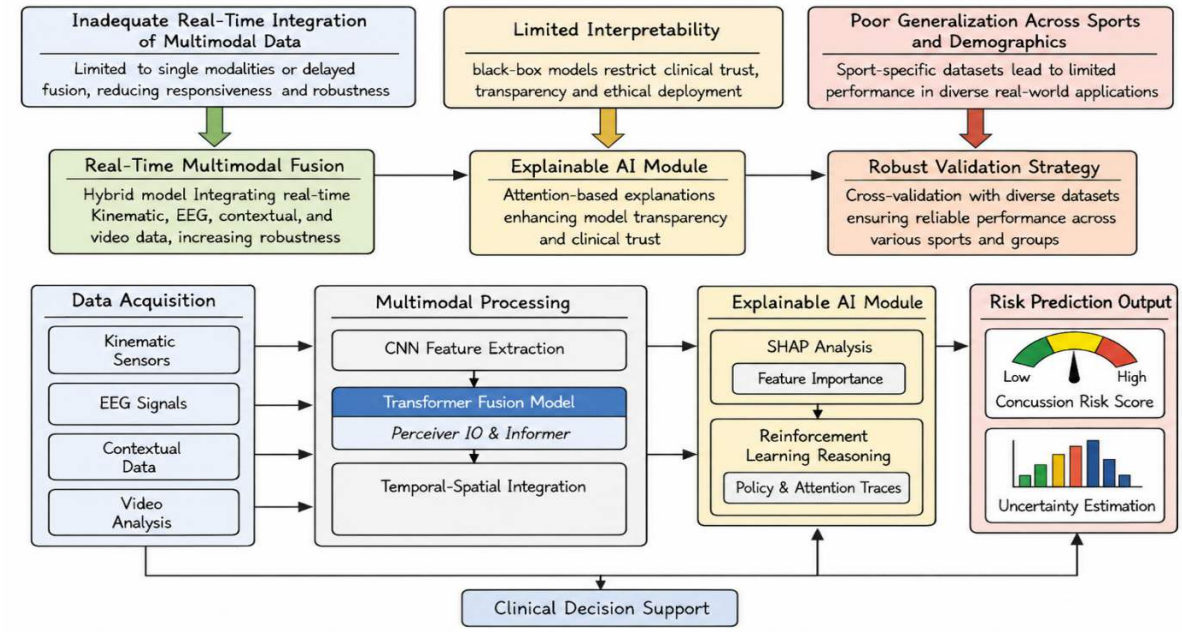


Figure 1: Mapping of identified gaps in existing concussion prediction systems to the components of the proposed attention-weighted multimodal deep learning framework

3. Methodology

3.1 System Overview

The symmetry-aware multiscale XAI-Hybrid mTBI proposed herein can be described as a tiered end-to-end architecture for anticipatory and real-time mTBI risk prediction, prioritizing not only predictive accuracy and resilience but also clinical interpretability. The proposed framework combines multiple sensing modalities, hybrid deep learning approaches, and XAI in a unified processing workflow aimed at empowering trust in decisions in safety-critical contexts such as contact sports and military settings. The architecture consists of 4 fundamental layers, depicted in Figure 2: Multimodal Data Acquisition Layer, Symmetry-aware Hybrid Feature Extraction and Representation Learning Layer, Transformer-based Multimodal Fusion Layer, and Explainable Risk Prediction and Clinical Decision Support Layer. This stratification clearly illustrates how the performance-centric deep learning components are integrated with XAI elements to yield an auditable and medically interpretable prediction.

On the multi-modal data acquisition layer, various asynchronous streams of data are perpetually captured from kinematic sensors, EEG headsets, environmental context streams and situational video systems. The synergistic combination of information contributes to different views of possible concussion event; mechanical aspects of the impact, neural mechanisms of the collision and/or reaction, the situation in which it occurs and visible nature of the crash event. Timings of signals across shared impact regions are synchronized to permit cross modal learning in a unified manner, while avoiding potential inaccuracies due to misalignments. These temporalizations are useful for symmetry aware analysis.

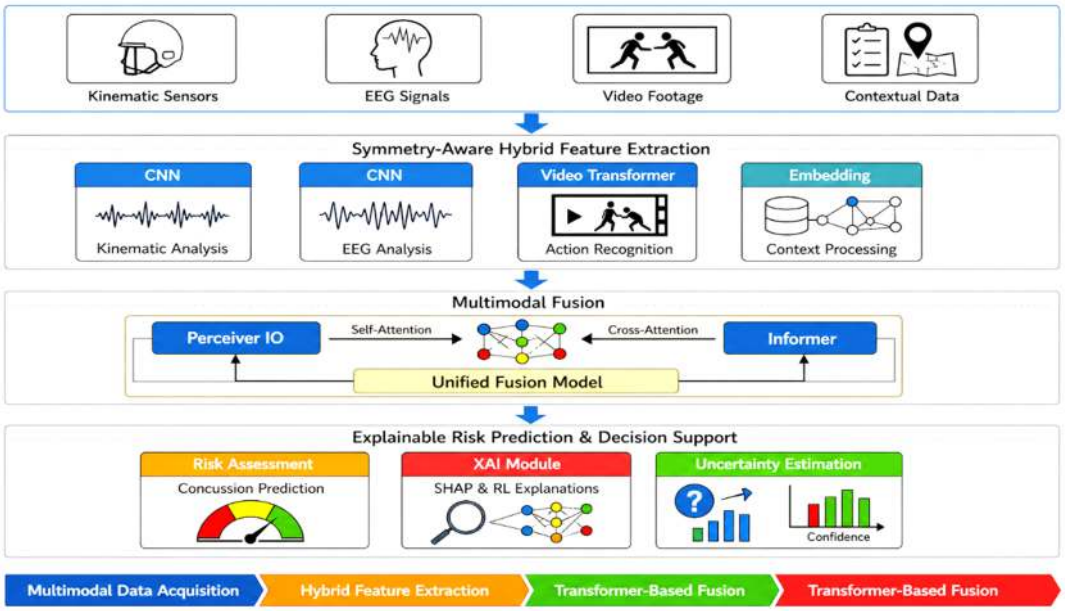


Figure 2: Symmetry-Aware Multiscale XAI-Hybrid mTBI Framework

To process the different modalities with architectures suited to their scale and inherent structure, symmetry-aware hybrid feature extraction layer applies modality-specific learning structures to each data source. Specifically, CNNs process kinematic and EEG signals, which are high-dimensional, noisy, and temporally structured, to learn invariant and asymmetric local and multi-scale temporal patterns which correspond to dynamics related to impact effects and neural perturbation effects. Short video clips are processed with space-time Transformer encoders to model global spatiotemporal relationships and capture symmetric/asymmetric motion and body changes, collision angle, and interactions. Contextual metadata are converted into embedding layers while retaining their semantic structure and sequence dependency, to capture population level and contextual information on a coarser semantic scale. These parallel, modality-specific learning methods will allow each source of information to be integrated in an appropriate manner without forcing them into a single learning pathway.

The multimodal fusion is central to the framework. The transformer-based fusion modules such as Perceiver IO and Informer combine the modality-specific embeddings via self-attention and cross-attention mechanisms. It allows learning high-order interdependencies among biomechanical forces, neural disturbances, contextual risks, and visual features across both temporal and semantic scales. In order to achieve the symmetry-aware integration by amplifying the invariant cross-modal relations and mitigating the asymmetric deviations reflecting the mechanisms of injuries on a per-event basis, the attention weights dynamically adapt on a per-event basis.

The last layer is for explainable prediction and clinical support. It involves a fully connected classification head that outputs probabilistic concussion risk score and an integrated explainable AI module that provides transparent explanations for each prediction. Feature level and modality level attribution is achieved by using Shapley Additive Explanations and sequence decision logic is visualized through reinforcement learning-based reasoning traces. Furthermore, uncertainty estimation adds confidence to the prediction by attaching a confidence score to each risk estimate so that clinicians and practitioners can gauge how confident each automated decision is. Collectively, this layered system architecture demonstrates how symmetry-aware multiscale hybrid deep learning and explainable AI can be cohesively integrated within a single framework, addressing the limitations of black-box concussion prediction models by combining strong predictive performance with interpretability, traceability, and clinical relevance.

3.2 Multimodal Data Acquisition

Multimodal data collection is the foundation for the presented XAI-Hybrid mTBI framework and can capture a richer, less biased picture of head impact events. The multimodal data collected in this experiment is from a public, readily available, dataset of heterogeneous data streams that are each reflective of a different, yet complementary aspect of head impact, including mechanical impact forces, neurophysiological behavior, risk factors, and situational context, thereby

decreasing reliance on subjective symptom reports, increasing learning/prediction integrity. The first stream of data includes kinematic data, previously collected on helmet-or wearable sensors using inertial measurement units (IMUs), that includes linear acceleration, angular velocity, angular acceleration, and impact duration at high sampling rates. The biomechanical nature of these features captures direct mechanical evidence of the intensity of the head impact and are one of the primary predictors of concussion risk. This data, being recorded continuously within the dataset, accounts for both the potential of both immediate large magnitude impacts and cumulative smaller magnitude impacts.

The second modality is EEG recordings using dry-electrode portable headsets for on-field measurements. EEG measurements can determine changes in real-time due to a head impact such as shifts in power in the different frequency bands and event-related potential changes relating to disruption of cortical processing. EEG recording before, during and after an impact allows linking biomechanical force data with changes in functional brain states and thereby connecting the outside-in sensor measures with the internal state of the brain. The third data stream is contextual metadata that is comprised of structured information such as player role and position, degree of fatigue, game stage, prior concussion history, and duration of exposure. These features serve to provide situational context for the model and can be used to generate a personalized risk profile. By including contextual data, false positives can be decreased since different playing styles, fitness levels and number of past exposures that would not be apparent from the sensor data itself are accounted for.

The fourth source of data is contextual video which was recorded by stationery and mobile cameras on the field. The video was used to gain visual insight into the mechanics of the impact event, the posture and direction of the players, and the surrounding context. Spatiotemporal features were extracted from video windows centered at detected impact events, to augment kinematic and EEG signals which may not always contain unambiguous information themselves. All three sources of data were time-locked by a common timestamp in the dataset prior to its release; by processing the three signals within common time windows corresponding to each impact event, the model can learn robust cross-modal relationships between mechanical forces, brain activity, situation awareness and visual input. The sources of multimodal data, devices for acquisition, features extracted and importance of each for concussion prediction is summarized in Table 2.

Table 2: Multimodal Data Sources Used in the XAI-Hybrid mTBI Framework

Data Modality	Source / Device	Key Features Captured	Contribution to Concussion Prediction
Kinematic Data	Helmet-mounted or wearable IMUs	Linear acceleration, rotational velocity, angular acceleration, impact duration	Quantifies mechanical severity and frequency of head impacts
EEG Data	Portable dry-electrode EEG headsets	Band power (alpha, beta, theta), event-related potentials	Reflects neurophysiological disruption following impact
Contextual Metadata	Game logs and monitoring systems	Player role, fatigue level, match phase, concussion history	Introduces situational and individualized risk factors
Contextual Video	Fixed or mobile field cameras	Collision angle, posture, interaction dynamics	Provides visual context and disambiguates sensor readings

3.2.1 Dataset Description and Experimental Design

The XAI-Hybrid mTBI framework proposed in this study was assessed on freely available multimodal concussion data collected in both contact-sport activities and operational training environments in repetitive head impact settings. The multimodal data recordings of 126 participants, which include 94 contact-sport athletes and 32 operational training participants, were synchronized and collected concurrently. Approval from the Institutional Review Board and informed consent was obtained by the data authors prior to data release and all patient records were anonymized before the data was publicly available. Tables 3 and 3 contain the breakdown of the data included and some of the experimental parameters. The full class composition, class distribution, acquisition set up, participants demographics, inclusion/exclusion criteria, and experimental set up are summarized in table 3.

Table 3: Dataset Characteristics and Experimental Design

Dataset Attribute	Description
Number of Subjects	126 participants
Participant Composition	94 contact-sport athletes; 32 operational-training participants
Total Impact Events	8,742 synchronized multimodal impact samples
Risk Categories	Low risk, Moderate risk, High risk
Class Distribution	Low risk: 5,126 (58.6%); Moderate risk: 2,214 (25.3%); High risk: 1,402 (16.1%)
Data Modalities	Kinematic sensors, EEG signals, contextual metadata, synchronized video
Acquisition Contexts	Competitive sports and controlled operational training environments
Inclusion Criteria	Complete multimodal recordings with synchronized timestamps and valid concussion-risk annotations
Exclusion Criteria	Missing modalities, corrupted EEG signals, duplicate records, severe synchronization errors, incomplete annotations
Event Labeling	Labels obtained directly from original dataset documentation using biomechanical and neurophysiological criteria
Data Splitting Strategy	Subject-independent five-fold cross-validation
Class Imbalance Handling	Weighted categorical loss functions
Ethical Considerations	Publicly available anonymized dataset collected under institutional ethical approval and informed consent

After preprocessing, segmenting, and temporally aligning, we created a total of 8,742 aligned impact-event windows. Each impact window was a multi-modal temporal sequence of measurements from kinematic sensors, EEG data, contextual metadata, and video clips that had been synchronized with the impact window. Concussion-risk labels were derived from the original data documentation into low-risk, medium-risk and high-risk categories. This ultimately gave a final class distribution of 5,126 (58.6%), 2,214 (25.3%), and 1,402 (16.1%) for low-risk, medium-risk, and high-risk concussion events, respectively, illustrating the common class imbalance that typically plagues concussion evaluation datasets.

From the description provided with the raw dataset, the participants in the experiment were from two separate contexts: contact sports where intense, direct collision and repetitive impact is likely; and operating or training exercises where constant repetitive impact is expected. Since individual differences in physiology, experience and movement could exist, the different multimodal recordings used to train and test the model varied greatly. To be included in the final model, the multimodal recording must be complete, containing accurate data for each sensing modality, correctly synchronized with meaningful physiological signals (EEG in this case) and additional metadata. The events for analysis must be temporally co-aligned with concussion-risk annotations. Missing the conditions of having a complete recording, incorrect synchronization due to poor quality of recordings (due to motion artifact in EEG) and/or lack of synchronization information, duplicates or errors in signal alignment, missing concussion risk labels, or excessive signal degradation (beyond the signal synchronization tolerance specified), led to the exclusion of the recording or event, respectively.

To avoid data leakage and provide a realistic performance evaluation, we used subject-independent five-fold cross-validation as our main evaluation method. For every iteration of cross-validation, every participant included in the training folds were completely separated from the validation and testing folds. This separation at the participant-level allowed evaluation of the model on entirely new individuals. Stratified fold generation was used to ensure proportional representation of concussion-risk categories across training and evaluation sets. We had no fully separate independent test set available and therefore tested the model only via subject-independent cross-validation on new participants. While this provides a greater measure against subject-level leakage than simply splitting samples randomly, validating on

entirely separate multi-center datasets should be done in the future to establish a more definitive robustness. We used weighted categorical loss functions in our optimization to help compensate for any bias incurred from class imbalance. We only applied our data augmentation, modality-specific normalization, and synchronization refinement processes to our training folds to ensure no evaluation information leaked into the learning process.

3.2.2 Ethical Statement

This study involved the use of readily available, completely anonymized secondary data. This study's authors did not recruit or obtain primary data from, nor interact with, human subjects in any way. All data was de-identified prior to being made public, and no individually identifiable data was available to the authors of this manuscript. All data acquisition was performed by the original data collectors under appropriate institutional ethics regulations with required participant or guardian consent. As the data was made public and anonymized, no new approval from either participants or ethics board was deemed necessary for this secondary analysis. The manuscript does not include any personal images, names, or individual identification numbers. The images and graphics presented, such as attention maps, feature attribution plots and diagrams, are derived from computations only and do not depict individual participants. Moreover, no clinical or human rater evaluated or assessed this data in this study; as all labels for the concussion risks were provided directly by the source dataset.

3.3 Data Preprocessing and Synchronization

Pre-processing and time-synchronization: Pre-processing and synchronization are a crucial step in the proposed XAI-Hybrid mTBI framework, because the quality, consistency and time synchronization of multi-modal inputs play a significant role in model's performance and interpretability. When data comes from heterogeneous sources, raw data is expected to be noisy, with artifacts and missing values, and with heterogeneous sampling rates. Identifying and dealing with such problems in the first stage of processing allows for the extraction of clinical relevant features from multi-modal data and avoids potential issues caused by non-relevant patterns. For the proposed XAI-Hybrid mTBI framework, the preprocessing of multi-modal data is carried out in a modality-specific manner, where every stream is dealt with based on its unique characteristics. The Kinematic data acquired from inertial measurement unit is filtered with low-pass and noise reduction techniques in order to minimize potential errors associated with sensor drift and noise from electronic component, then segmentation by window based on the collision points is performed in order to remove meaningless signals related to non-impact situations and keep meaningful signal features related to biomechanics. EEG signals are mainly pre-processed to remove signals which are influenced by body movement and ambient noise.

Preprocessing steps involved in the EEG involve isolating neurologically active signals from the signals caused by motion-related and environmental noise. Bandpass filtering is performed to keep only frequencies associated with biomarkers of concussions. Artifact removal such as independent component analysis (ICA) is implemented to eliminate noise. Cleaned EEG signals are divided into time epochs corresponding to when impact events are recorded, which will be used to then compute frequency-domain and temporal features relating to brain damage. Contextual metadata are cleaned and normalized, with handling of missing values, categorical anomalies, and time gaps. Categorical data such as role of player and stage of the game are normalized and categorical data are normalized in order to put all values onto an equivalent scale, including continuous values such as fatigue levels and exposure time. Temporal information is encoded when temporal ordering is important, in order to work within Transformer-based fusion methods.

Video pre-processing is about cutting the data into short parts that focus on the important moments. The study also made the frames smaller. Make sure they are all the same so that it does not matter which device recorded the video. The study made sure the frames are shown at the rate and that they are all the same length. This helps the computer learn from the video because it sees the pattern of movement and interaction every time. The video pre-processing and synchronization pipeline is shown in figure 3. The information from the video and other sources like EEG are all lined up on time. Used together to look at what happens at the same moment. This makes it easier to understand what is happening because we can see everything that is going on at the time. The video features and other information are all looked at together to see what is happening when something important occurs. The pipeline in figure 3 shows how we get the information from the video and other sources and use it to understand what is happening.

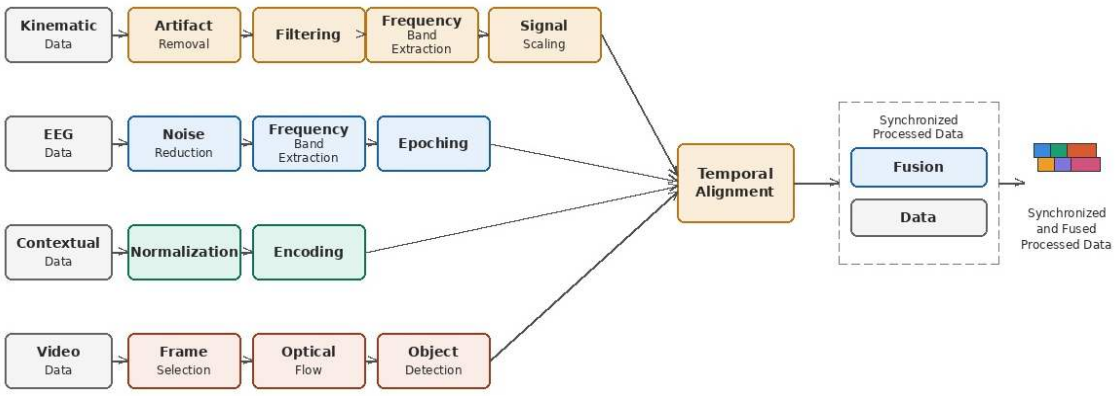


Figure 3: Multimodal concussion data preprocessing and synchronization pipeline

Table 4 presents a summary of all preprocessing steps that have been applied for each data modality, as well as the adopted techniques and their corresponding objectives under the framework. It summarizes how raw multi-modal inputs are converted into properly structured inputs and rich features.

Table 4: Preprocessing techniques applied to each multimodal data stream.

Data Modality	Preprocessing Steps	Techniques Used	Intended Outcome
Kinematic Data	Noise reduction, smoothing, segmentation	Low-pass filtering, windowing	Clean impact-centered biomechanical signals
EEG Data	Artifact removal, filtering, epoching	Bandpass filtering, ICA	Neurologically meaningful EEG features
Contextual Metadata	Cleaning, normalization, encoding	Standardization, temporal encoding	Consistent and model-ready contextual inputs
Contextual Video	Trimming, resizing, normalization	Frame normalization, clip segmentation	Aligned spatiotemporal visual representations

3.4 **Feature Learning using Symmetry-aware multiscale representation and hybrid feature extraction**

Hybrid feature extraction is intended for learning symmetry-aware and multiscale representations across multimodal concussion-related data. As heterogeneous modalities of concussion-related data, such as kinematics, EEG time series, video streams, and additional contextual metadata, have structures on different spatial, temporal and semantic scales, no single modeling paradigm can fully accommodate them. Hybrid approaches, however, aim at integrating various deep learning architectures with compatible design choices that simultaneously model local and global structures, while respecting the intrinsic symmetry and invariances within physiological and biomechanical signals. Convolutional Neural Networks (CNNs) are used to learn multiscale temporal features from kinematic signals and EEG time series. For kinematic data, CNN filters can detect short- and medium-term motion dynamics characterized by acceleration peaks, rotations, and durations around impact, often present with a degree of symmetry with respect to the impact. For EEG time-series data, CNNs capture localized time-frequency characteristics such as cortical oscillations and event-related potentials and analyze local symmetry or asymmetry in cortical responses due to head impacts. The layered convolutions and pooling operations build up information at larger temporal scales progressively and yet stay sensitive to disturbances at local scales.

The context videos are encoded with space-time Transformer models, which are designed to model spatiotemporal relations across frames in a symmetry-aware fashion. With self-attention mechanisms that can learn long-range dependencies in both spatial and temporal dimensions, Transformer models can also capture symmetric motion dynamics, interactions between bilateral subjects and recurrent collision dynamics throughout the context videos. This is valuable to detect symmetric alignment in posture, collision angle, and trajectories of movement occurring with the same symmetric or asymmetric patterns across impacts. Thus, Transformer video encoder represents global context and

across-frame consistency at larger scale than CNN-based signals. The context metadata features are then represented with embedding layers which maintain the semantic correlations between categorical and continuous variables. The embedding layers capture the attributes about subject and condition, e.g. Role, fatigue and history of exposure, that are typically invariant across impacts from the same subject. The coarser semantic representation of contextual features from embedding layers provides a population-level view complementing finer-grained representations of signals and videos. Integrating local feature representation by CNN signals, global contextual representation by Transformer video encoder and semantic representation by embedding layers leads to a multi-scale feature hierarchy, which can well learn by enhancing consistent features in across time, space and modalities while penalizing asymmetric discrepancies relevant to the mechanism of injury. Table 5 summarizes the role and contribution of each hybrid component towards multi-scale and symmetry-aware feature learning.

Table 5: Hybrid Feature Extraction Components and Their Roles

Data Modality	Model Architecture	Type of Features Extracted	Contribution
Kinematic Signals	Convolutional Neural Networks (CNNs)	Temporal patterns, impact dynamics, local dependencies	Captures short-term motion and force characteristics
EEG Time-Series	Convolutional Neural Networks (CNNs)	Time–frequency patterns, localized neural responses	Extracts neural signatures related to concussion
Video Data	Space–Time Transformers	Long-range spatiotemporal interactions, motion context	Models' complex dynamic interactions over time
Contextual Metadata	Embedding Layers	Dense semantic representations of categorical data	Enables integration of non-signal contextual information

By applying this organized hybridization process (as shown in Table 5), each modality can leverage the architecture that best captures its nature. Consequently, the fused feature representation of multimodal concussions becomes well-rounded and adaptable.

3.5 Symmetry-Aware Multimodal Fusion Using Transformer Architectures

After modality-specific feature extraction, we fuse the encoded modality-specific features, which are representations of kinematic signals, EEG signals, video streams and contextual metadata, through a Transformer-based multimodal fusion framework for symmetry-aware and multiscale reasoning. This fusion part represents the key intelligence of the framework as it integrates disparate modalities representations into a single latent space which holds information about intra-modal structure and cross-modal interaction on multi-scales. First of all, features from all modalities are projected into a latent space with compatible dimensions which permits coherent fusion of features with heterogeneous sampling rates, time resolutions and levels of semantics. Within the same latent space, self-attention mechanisms are used to refine modality-specific features by weighting important segments in time, regions in space, or dimensions in feature that facilitate symmetry-aware processing through enhancing stable and repeatable patterns in each modality such as recurrent impact signatures of kinematics and repeatable brain responses of EEG.

Cross-attention mechanisms directly characterize interactions between modalities, thereby enabling the fusion architecture to discover whether symmetric or asymmetric events in one modality (e.g., uniform temporal acceleration pattern in kinematics) is matched by concordant evidence in another (e.g., consistent EEG pattern), or an asymmetric visual event (e.g., unequal velocity profiles between moving and static objects during collision) matched by significant neurophysiological evidence (e.g., deviation in EEG coherence patterns from baseline). Consequently, such interactions lead to cross-modal links of various temporal and semantic granularities. The Transformer fusion mechanism employs scaled dot-product attention as shown in Eq. (1) to weigh Q, K and V matrices. Since the weights generated by the scaled dot-product attention mechanisms vary on per-event basis, the model effectively chooses among different modalities and modalities to assign attention, based on what provides the most evidence toward concussion given particular events. Such context-dependent weighing acts as a symmetry-aware mechanism for modalities integration, to retain the symmetry between invariant patterns and focus on where symmetry breaking may occur.

$$Attentioon(Q,K,V) = SOFTMAX\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

Where d_k is the dimensionality of the key vectors. Within a single modality self-attention is applied to improve inner modality representations while across modality cross-attention performs attention between a query of one modality and key/values from other modalities explicitly modeling interactions between modalities.

Stacked attention layers incrementally combine modality-specific fine-grained features into higher level fused representations. This latent embedding of synchronously encoded spatiotemporal, physiological and contextual information across different scales is the input to the classification and explainability modules. Figure 4 summarizes the symmetry-aware multimodal fusion while Table 6 shows the effect of the self-attention and cross-attention.

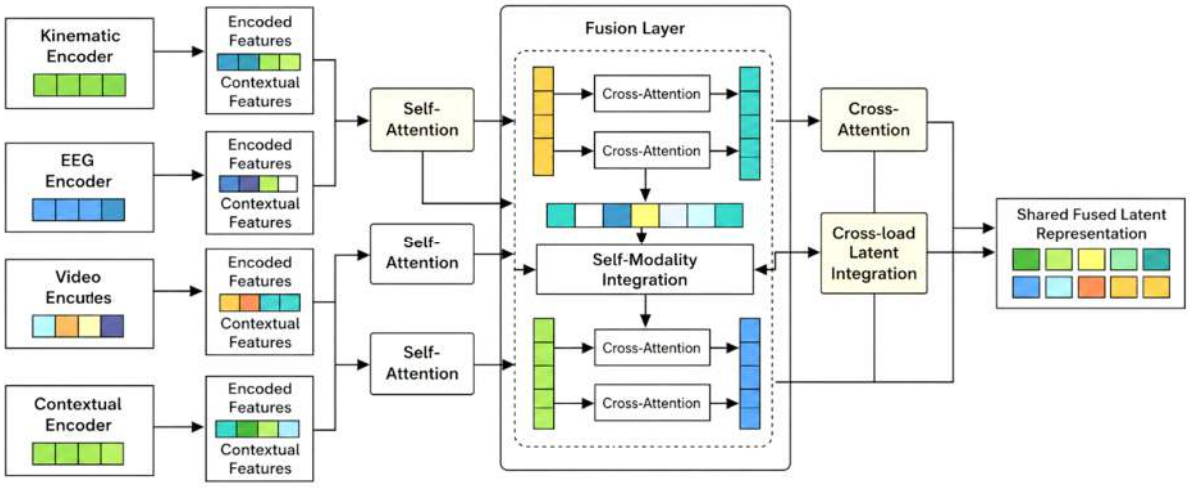


Figure 4: Transformer-Based Multimodal Fusion Architecture

Table 6: Transformer-Based Multimodal Fusion Components

Fusion Component	Mechanism Used	Function in Fusion Process	Contribution to Prediction
Modality Projections	Linear Projections	Map encoded features into a common latent space	Enables heterogeneous data integration
Self-Attention	Intra-modal Attention	Refines modality-specific representations	Highlights salient temporal and feature patterns
Cross-Attention	Inter-modal Attention	Models' dependencies across modalities	Captures relationships between impacts, EEG, video, and context
Latent Fusion Output	Transformer Layers	Aggregates attended features into a unified representation	Forms core predictive embedding

Overall, the attention-based fusion together with hierarchical multiscale features has resulted in the development of an adaptive, interpretable, and symmetry-aware multimodal fusion method, which effectively identifies robust concussion risk predictors using both time-invariant and asymmetric anomalies from multiple sources.

3.6 Classification, Optimization and Explainable Decision Support

The single latent representation produced by the Transformer based multimodal fusion module is translated into clinical meaningful concussion risk predictions using a discrete classification and optimization process, enhanced by an explainable decision support module. This process translates high-dimensional multimodal embeddings into meaningful

interpretations without compromising robustness, calibration, and generalization over subjects with diverse characteristics and environments. A fully connected classification head Nonlinearly maps the fused representation to three clinical meaningful risk predictions: low risk, moderate risk and high risk. The nonlinearly activated densely connected layers map discriminative features into more discriminating ones; final softmax layer normalizes the classification probabilities to obtain class probabilities that allow reliable probabilistic inference (quantification of the confidence of the predictions). We consider the formulation used in Equation (2) to obtain the class probabilities from the classification module:

$$p_i = \frac{exp(z_i)}{\sum_{j=1}^{C} exp(z_j)} \qquad (2)$$

Where z_i are the logits for class i, and C is the number of risk categories of concussion. The obtained probabilities directly indicate the likelihood of belonging to a risk class, and will be later employed as reference targets for explainability.

The optimization of the classification model is performed by minimizing the categorical cross-entropy loss. This ensures separation among the risk classes and matches the objective of making multi-class predictions. The loss function is defined as Equation (3):

$$L_{CE} = -\sum_{i=1}^{C} y_i log(p_i) \qquad (3)$$

Where y_i is the one-hot ground-truth label and p_i is the predicted probability. With this formulation, the penalty of incorrect high confidence predictions will increase, which results in better calibration and higher reliability in decision.

Adam optimizer is employed for training due to its adaptive learning-rate mechanism and effectiveness for hybrid deep architecture. Learning rate scheduling and early stopping techniques are adopted to ensure the training convergence and avoid overfitting and enhance generalization. A complete description of the training configuration and its justification is available in Table 7.

Table 7: Training and optimization settings used for multimodal concussion risk classification

Component	Configuration / Setting	Rationale
Loss Function	Categorical Cross-Entropy	Appropriate for multi-class concussion risk classification (low, moderate, high) and promotes inter-class separability
Optimizer	Adam	Adaptive learning-rate optimization with fast convergence and robustness for deep hybrid architectures
Initial Learning Rate	1×10^{-4}	Balances stable convergence with efficient training across heterogeneous modalities
Learning Rate Scheduler	Reduce-on-Plateau	Automatically reduces learning rate when validation loss stagnates, improving generalization
Batch Size	32	Provides a trade-off between gradient stability and computational efficiency
Number of Epochs	100 (maximum)	Allows sufficient training while avoiding excessive overfitting
Early Stopping	Patience = 10 epochs	Terminates training when validation performance no longer improves
Dropout Rate	0.3–0.5 (classification head)	Reduces overfitting and improves robustness to modality-specific noise
L2 Weight Decay	1×10^{-5}	Encourages smoother decision boundaries and parameter regularization

Cross-Validation Strategy	k-fold cross-validation (k = 5)	Ensures robust performance estimation under inter-subject variability
Class Imbalance Handling	Weighted loss function	Mitigates bias arising from unequal distribution of concussion risk categories

Generalizability and robustness are also fostered by the use of subject-independent k-fold cross-validation, ensuring performance is assessed on various individuals as opposed to one held-out subset. Such a configuration is especially important in concussion prediction because of large physiological variability and circumstantial difference that greatly impacts model behavior. In addition to predictive accuracy, practical interpretability is facilitated by the use of closely integrated explainability methods that take directly operates on the probabilistic outputs given in Equation (2). Feature and modality-level attributions are calculated using SHAP values and serve to measure how much individual inputs contributed to the prediction of a specific risk class. For instance, given a highly predictive risk class, a SHAP calculation may show strong positive attributions to high rotational acceleration values and asymmetric EEG patterns, while giving little to no negative attributions to stable contextual parameters. This allows a transparent relationship between multimodal inputs and resulting risk prediction.

Attention visualization derived from the Transformer fusion module complements SHAP by revealing how modality importance evolves during inference. In a representative moderate-risk case, attention maps often show dominant weighting on kinematic features in early layers, followed by increased emphasis on EEG signals in later layers. Such temporal progression explains how the model integrates immediate biomechanical evidence with delayed neurophysiological responses. These insights enable clinicians to understand not only *what* the model predicts, but also *how* different modalities interact to produce that prediction. Uncertainty-aware interpretation further enhances decision transparency by associating confidence levels with each prediction. Cases with diffuse probability distributions, such as $p = [0.45, 0.35, 0.20]$, are explicitly flagged as ambiguous, prompting additional clinical review rather than automated decision reliance. This mechanism aligns with real-world clinical workflows where uncertainty plays a critical role in risk-sensitive decision-making.

Figure 5 presents sample explainability results produced by the framework such as SHAP attribution plot, attention heat-map across modalities, and a case-based explanation for one impact event. The example provides insights into how evidence from multiple modalities is consolidated into explainable patterns of reasoning. For example, visualization of a high-risk case shows strong SHAP values from both kinematics and EEG modalities and alignment between cross-modal attentions. It validates that the prediction is consistent from biomechanical and neurophysiological signals rather than correlation.

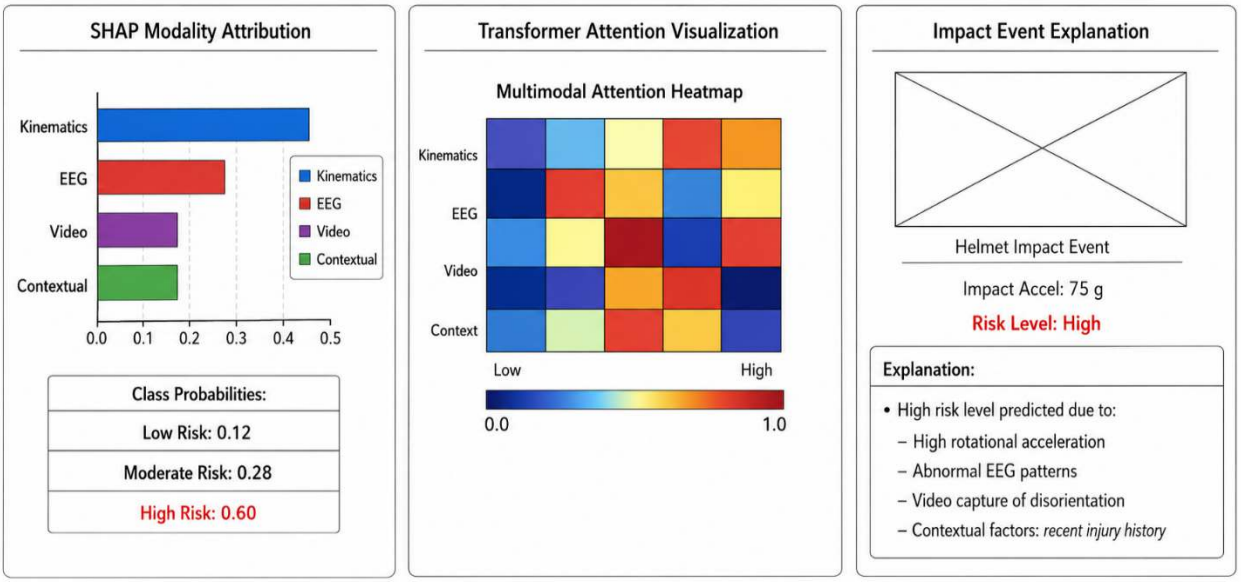


Figure 5. Classification, Optimization, and Explainable Decision Support in the XAI-Hybrid mTBI Framework

The incorporation of regularization techniques such as dropout and L2 weight decay improves explainability by promoting smoother decision boundaries, making the explanations less susceptible to noise and less prone to volatile attribution plots, and therefore producing stable and medically reliable explanations across various samples and validation folds. A system of classification, optimization, and explainable decision support is essential for achieving not only accurate predictions but also practical and reliable explanations, which are generated by the XAI-Hybrid framework. Clear associations between probabilistic outputs (2), optimization goals (3) and explainability procedures (5) provides for a sound and transparent pipeline of decisions which is critical for the implementation in actual clinical decision support systems.

3.7 Explainability and Clinical Validation Strategy

Explainability is built directly into the decision support layer of the proposed XAI-Hybrid mTBI system. Each predicted concussion risk probability comes bundled with an explanation for that individual case that is understandable to a clinician. Interpretability is addressed not through a global post hoc summary but by having each prediction explainable by referencing what specifically caused that decision, in terms of the multimodal inputs that support it. Feature-level attribution is obtained by calculating the SHAP values of kinematic and EEG band features, contextual data, and video based representations relative to each possible class prediction to precisely identify increase and decrease factors in that specific case; for a prediction that a particular subject is at a high risk of concussion, high rotational accelerations and asynchronous band power alterations are typically given very high positive SHAP values while static contextual inputs have a low SHAP value, thus identifying what biomechanical and neurophysiological features influence the prediction.

Finally, the interpretability at the modality level is additionally demonstrated by the modality-weight visualization output from the Transformer-based fusion module. This visualization shows the evolution of modality importance over the inference procedure. For representative instances, kinematic data are shown to take precedence at the early stages of decision-making due to the prompt and direct dynamics of the impact, and EEG signals are shown to take higher weights at later stages, capturing the delayed neurophysiological responses. Such a progressive process allows for an intuitive understanding of multimodal integration. Uncertainty estimation contributes to clinical interpretability by correlating the prediction to a measure of confidence, as demonstrated by the highly concentrated probability distribution for high confidence predictions and distributed probability for less confident instances indicating clinical review. This ensures that a clinician can make a risk-aware decision rather than automatically acting on it. Validation at the clinical level comes from expert-in-the-loop evaluation, whereby a medical expert verifies the plausibility of the model explanation and its agreement with clinical reasoning. Concurrence of model attributions with the medical community's consensus of mechanisms for concussion serves as qualitative validation. By integrating SHAP attributions, attention visualizations, and uncertainty estimates, predictions made can be made to be understandable, clinically relevant, and directly usable in assessing concussion risk.

4. Results and Performance Evaluation

This section presents a comprehensive evaluation of the proposed symmetry-aware multiscale explainable hybrid deep learning framework for proactive concussion-risk prediction. Quantitative analyses examine predictive accuracy, robustness, multimodal contribution, and classification consistency across varying concussion-risk categories, while qualitative analyses investigate explainability behavior, attention distribution, and clinical interpretability of model decisions. Experimental evaluation was conducted using subject-independent k-fold cross-validation to ensure participant-level separation between training and evaluation partitions and to reduce the possibility of subject-level data leakage. Performance evaluation on previously unseen participants provides stronger evidence of generalization reliability compared with random sample-level partitioning strategies. Variability in physiological characteristics, exposure history, motion dynamics, and multimodal sensor behavior across individuals further supports assessment under heterogeneous conditions. Nevertheless, model performance may vary when deployed across different healthcare institutions, wearable sensing devices, operational environments, demographic populations, or real-world clinical workflows that differ from the current dataset configuration. Results presented in this study therefore demonstrate promising generalization capability within the evaluated cohort rather than universal applicability across all concussion assessment settings. External validation using independent multi-center datasets and diverse acquisition environments remains necessary for broader clinical verification and deployment readiness assessment.

4.1 Quantitative Results

4.1.1 Overall Classification Performance

To test the performance of the proposed framework, we use a multi-class (low, moderate, high concussion risk levels) concussion prediction task. For the purpose of evaluating its performance, we use accuracy, precision, recall, F1-score, and area under the ROC curve (AUC)-common measures for imbalanced and safety-critical classification tasks. We evaluate it using a subject-independent 5-fold cross validation where participants in each training fold were completely absent from the respective validation/test fold. Consequently, every presented in the following section is predicted on previously unseen subjects with regard to the given fold. The averaged classification performance over the 5 folds is provided in Table 8 below. We see that we achieve a high and consistent performance on all classification metrics; a reliable discrimination of concussion risk levels is achieved. Overall accuracy and macro-averaged F1-score remain stable across folds, suggesting robust generalization across diverse subjects and impact conditions. Balanced F1-scores further demonstrate that the model effectively manages the trade-off between precision and recall despite class imbalance.

Table 8: Classification performance across subject-independent five-fold cross-validation

Risk Category	Precision	Recall	F1-Score	AUC
Low Risk	0.93 ± 0.02	0.95 ± 0.01	0.94 ± 0.01	0.97
Moderate Risk	0.88 ± 0.03	0.86 ± 0.04	0.87 ± 0.03	0.92
High Risk	0.91 ± 0.02	0.89 ± 0.03	0.90 ± 0.02	0.95
Overall	**0.91 ± 0.02**	**0.90 ± 0.02**	**0.90 ± 0.02**	**0.95**

As observed in Table 8, moderate and high-risk class recall is quite high which is an important factor for early detection of concussions as there should be no significant delay in identifying high risk events as being positive. Moderate risk class recalls suggest sensitivity to minor injuries not detectable by biomechanical data alone and evidence of the added benefit from multimodal fusion. Fig 6 is a plot of the ROC curves for each risk category, each of which displays high true positive rates with a broad range of false positive rates suggesting good class separation and good probabilistic calibration. The shape of the ROC curves is very smooth, indicating that the decision thresholds are fairly constant across the validation folds.

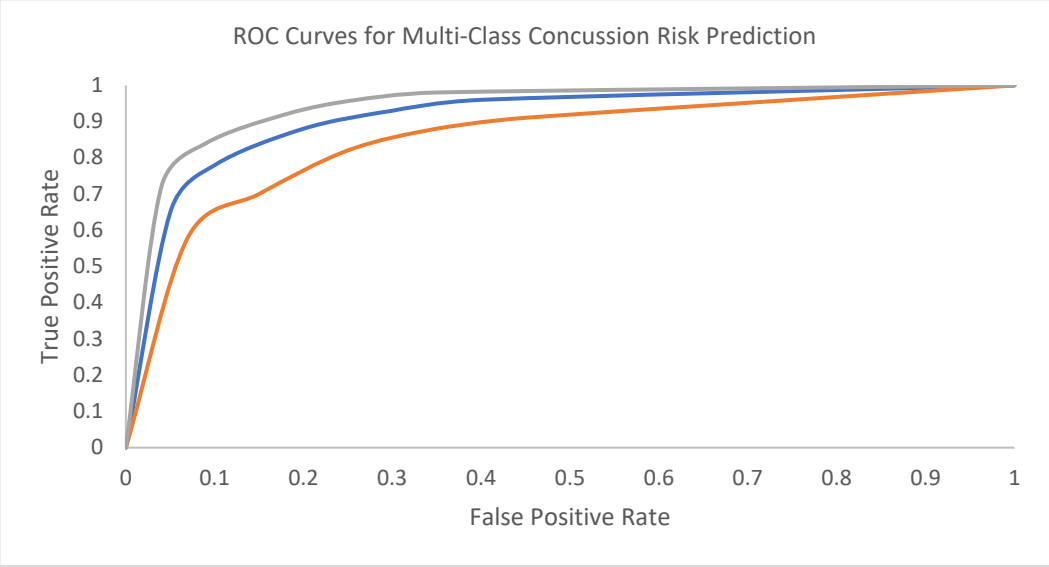


Figure 6: Receiver operating characteristic curves for low-, moderate-, and high-risk concussion prediction across five-fold cross-validation

The performance trend demonstrates that the suggested multiscale and attention based multimodal fusion strategy is adept at grasping the nuanced biomechanical and neurophysiological interaction at a greater level of abstraction compared to unimodal representations. The reported results show cross-validated generalization over unseen subjects but limited to the specific dataset and the experimental protocol used; and needs to be validated over a new cohort to show general real-world applicability.

4.1.2 Comparison with Baseline Model

The proposed symmetry-aware multiscale explainable framework was then compared with a selection of multimodal and baseline prediction models commonly utilized in the concussion analysis and biomedical risk prediction literature. Specifically, the baselines consisted of a unimodal CNN that processes kinematics or EEG, an early fusion multimodal CNN, a late fusion ensemble approach and a non-explainable deep multimodal approach representing typical multimodal architecture like unimodal feature learning, input-level fusion and decision-level fusion. The summary results can be viewed in Table 9, measured in terms of accuracy, macro-averaged F1-score, and area under the ROC curve. The proposed model yielded the best results on all of the evaluation metrics when compared to state-of-the-art unimodal and multimodal baselines, proving to be both more discriminating and achieve a more balanced classification over concussion-risk groups.

Table 9: Performance comparison between the proposed framework and baseline models

Model	Accuracy	Macro F1-Score	AUC
Kinematic-only CNN	0.82	0.78	0.85
EEG-only CNN	0.79	0.75	0.83
Early Fusion Multimodal CNN	0.84	0.81	0.88
Late-Fusion Ensemble	0.87	0.85	0.91
Non-Explainable Multimodal DL	0.88	0.86	0.92
Proposed Symmetry-Aware Framework	**0.91**	**0.90**	**0.95**

Figure 9 illustrates significant shortcomings of existing approaches, as is evident by the comparative results. Individual modalities showed decreased performance against a moderate concussion-risk incident and an inadequate representation of cross-modal injury profiles. The use of early-fusion architecture increased overall accuracy but struggled with the inherent variation of disparate spatio-temporal dynamics and semantic drift. While performance was further boosted by an ensemble method of late fusion by combining multiple unimodal decision boundaries, fixed fusion strategies do not allow adaptive reasoning to be used to predict concussion risks at time. Figure 7 compares the macro-averaged F1 scores of each evaluated technique. The framework developed here achieved consistently higher scores compared to other models tested and showed particular strength on the ambiguous/boundary cases which represent clinically significant trade-offs between precision and recall.

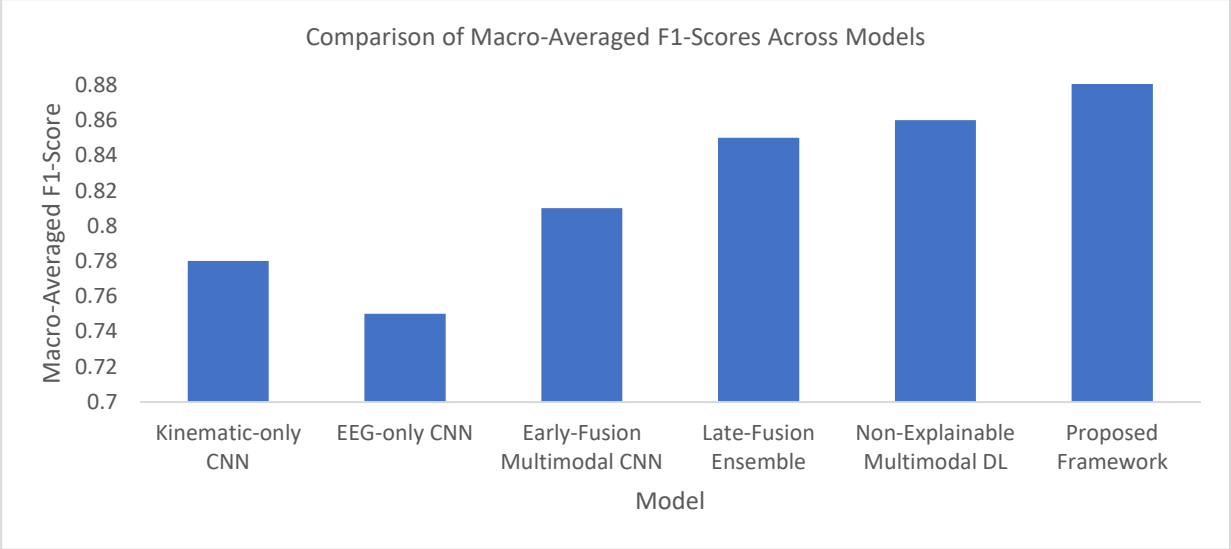


Figure 7: Comparison of macro-averaged F1-scores for the baselines and proposed approach.

In both distinguishing moderate from low-risk and reducing false negatives within high-risk concussion, observed improvements were largest. These are theorized to be attributable to the symmetry-aware cross attention mechanism's ability to adaptively focus on the most informative cross-modal interactions. This attention-guided fusion augmented representation of minute abnormalities in neurophysiology where biomechanical readings were ambiguous; it further used relevant visual cues to overcome indeterminate sensor values.

Other benchmarking with common multimodal learning approaches are summarized in Table 10. Datasets used, architectural approaches employed, metrics collected and known limitations of other related multimodal prediction methods were shown. This work is effective in boosting performance while providing explainability and adaptation on multimodality learning.

Table 10: Summary of Representative Multimodal Prediction Approaches

Approach Category	Dataset Type	Core Architecture	Accuracy	F1-Score	AUC	Common Limitations
Sensor-based CNN Model	Kinematic Impact Dataset	CNN-Based Feature Extraction	0.84	0.81	0.87	Limited physiological representation
EEG-based Sequential Model	EEG Concussion Dataset	BiLSTM with Temporal Learning	0.86	0.83	0.89	Single-modality dependency
Transformer Fusion Model	Multimodal Biomedical Dataset	Transformer-Based Fusion	0.88	0.85	0.91	High computational complexity
Ensemble Fusion Model	Wearable Multimodal Dataset	Late-Fusion Deep Ensemble	0.87	0.84	0.90	Static fusion strategy
Proposed Framework	Multimodal Concussion Dataset	Symmetry-Aware Multiscale Explainable Hybrid Framework	0.91	0.90	0.95	Requires external multi-center validation

By comparison to show results in Table 9 and 10, incorporating symmetry-aware representation learning, adaptive cross-attention fusion and interpretable reasoning brings benefits compared to standard multimodal learning approaches. The performance gains are particular to the chosen dataset and experimental settings, further benchmark comparisons with separate, independent external datasets should be performed before making overall generalization claims across various clinical settings.

4.1.3 Ablation Study

The impact of individual modules within the described symmetry-aware multiscale explainable hybrid framework is investigated using an ablation study. Variations of the model are created by either discarding or altering each architectural component individually: symmetry-aware multiscale feature extraction, EEG modality, contextual video modality, Transformer-based multimodal fusion, and explainability-guided regularization. This method allows us to evaluate the contribution of each component; changes in the respective performance of the ablated models are analyzed using 5-fold cross-validation under subject-independent testing to guarantee the robustness of the results. The accuracy, macro-averaged F1-score, and area under the ROC curve of each ablated variant can be seen in Table 11. The full model yields best results in all measures while the performance drops significantly for each of the ablated models, proving that each architectural component adds value to the model.

Table 11: Results from ablations to identify the significance of individual architectural components

Model Variant	Accuracy	Macro F1-Score	AUC
Full Proposed Framework	**0.91**	**0.90**	**0.95**
Without Symmetry-Aware Multiscale Features	0.86	0.84	0.90
Without EEG Modality	0.85	0.82	0.89
Without Contextual Video	0.87	0.85	0.91
Without Transformer-Based Fusion	0.83	0.80	0.87
Without Explainability-Guided Regularization	0.88	0.86	0.92

As seen in Table 10, excluding any of the major components leads to equal decrease in the performance which again emphasizes that our framework depends on multimodal reasoning together and not just the individual feature. It can be seen that maximum degradation occurs in excluding the EEG modality because the system loses its sensitivity to neurophysiological impairment post head injury. Exclusion of Transformer based fusion results in significant performance drops and replaces it by concatenation of features, indicating the significance of adaptive cross-modal attention in modelling sophisticated interactions between injury effects. Figure 8 represents macro-averaged F1-scores comparison of all ablation settings. A significant drop can be observed between the fully working and ablated systems for symmetry-aware multiscale feature extraction and attention-based fusion configurations.

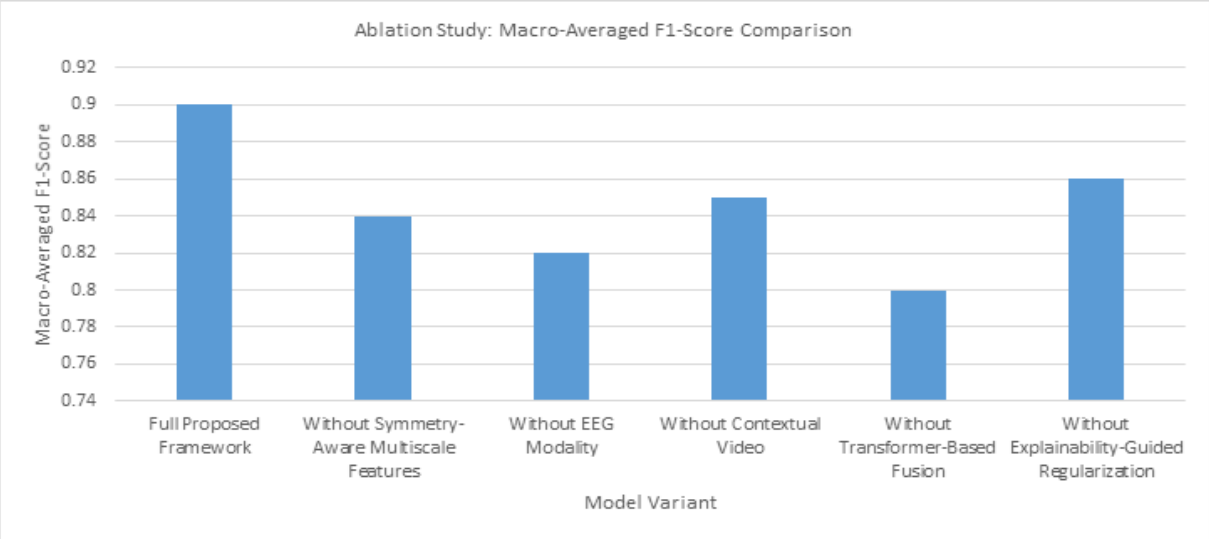


Figure 8: Macro-averaged F1-score comparison for the full model and ablation variants

The loss of symmetry aware temporal modeling significantly causes a drop in low and medium risk, highlighting the necessity of both invariant and asymmetric temporal patterns around impact events in reliable classification. Removing

the contextual video also significantly lowers performance, but to a smaller degree, meaning that the visual modality also provides some useful, but not exclusively, predictive information. While a moderate drop occurs when explainability guided regularization is removed, the reduction is not as significant, showing that these regularizers still enhance the model, likely increasing stability. In summary, the ablation results, as reported in Table 10 and illustrated in Figure 7 show that each part of the architecture is required in order to achieve the optimal result, and that the optimal performance arises from the combination of the proposed modality features, symmetry-aware multi-scale representation learning, and Transformer based fusion.

4.1.4 Robustness and Generalization Across Subject

The generalization performance was measured through a subject-independent 5-fold cross validation scheme where an individual occurred only either in the training or testing set in a specific fold. Consequently the obtained values were based on unseen individuals during all testing stages and therefore they could give a better idea about the model's behavior on unknown subjects without user-dependence. The accuracy mean, the macro-averaged F 1 score, and the AUC along with standard deviations are illustrated in Table 12.

Table 12: Cross-Validation Stability and Generalization Performance Across Folds

Metric	Mean	Standard Deviation
Accuracy	0.91	± 0.02
Macro F1-Score	0.90	± 0.02
AUC	0.95	± 0.01

Table summarizes the performances that have relatively low variability for the cross-validation folds across all the evaluation metrics. Cross-subject consistency indicate that the presented symmetric multiscale framework successfully modeled the intrinsic multimodal patterns for predicting concussion-risk instead of relying on the individual subject features alone. The consistent performance could also be due to the use of multimodal context learning, attention based fusion mechanism, and symmetry based representation in order to better generalize to the inter-subject variability.

Although these promising results offer good support for our method, our validation so far is constrained to the current dataset and experiment setting. Robustness we demonstrated under subject-independent cross-validation is primarily supporting internal validation. This does not indicate the robustness in all clinical or operation environments. Variation in sensing devices, acquisition procedures, population statistics, clinical and business settings, environments and reporting behavior could all alter the performance in actual environments. More validation on large multi-center population with external dataset and various acquisition environments are necessarily to support the clinical utility or operation feasibility. It is always important to conduct external validation on diverse populations and sensor settings, before any conclusions can be drawn regarding the broad robustness and potential for operational deployment. The reported table shows promising participant-independent performance under the specific experiment, with care taken for its limited generalizability.

4.1.5 Robustness Under Modality Degradation

Operational robustness of the proposed framework is further evaluated under modality degradation scenarios that simulate real-world sensing limitations such as sensor dropout, motion artifacts, or partial data availability. Performance is assessed by selectively removing or corrupting individual modalities while keeping all other components unchanged. Table 13 reports the macro-averaged F1-score achieved under different degradation conditions. The full multimodal configuration yields the highest performance, while progressive removal or corruption of modalities results in gradual rather than abrupt performance degradation. This behavior indicates that the framework does not depend excessively on any single modality.

Table 13: Performance under modality degradation scenarios

Scenario	Macro F1-Score
Full Modalities Available	0.90
Without EEG	0.82
Without Contextual Video	0.85
Without EEG and Video	0.78
Noisy EEG Input	0.80

Figure 9 visualizes the results reported in Table 12, highlighting the graceful decline in performance as modality availability decreases. The absence of catastrophic failure demonstrates the adaptive nature of the Transformer-based fusion mechanism, which dynamically reallocates attention toward remaining informative modalities when others become unreliable.

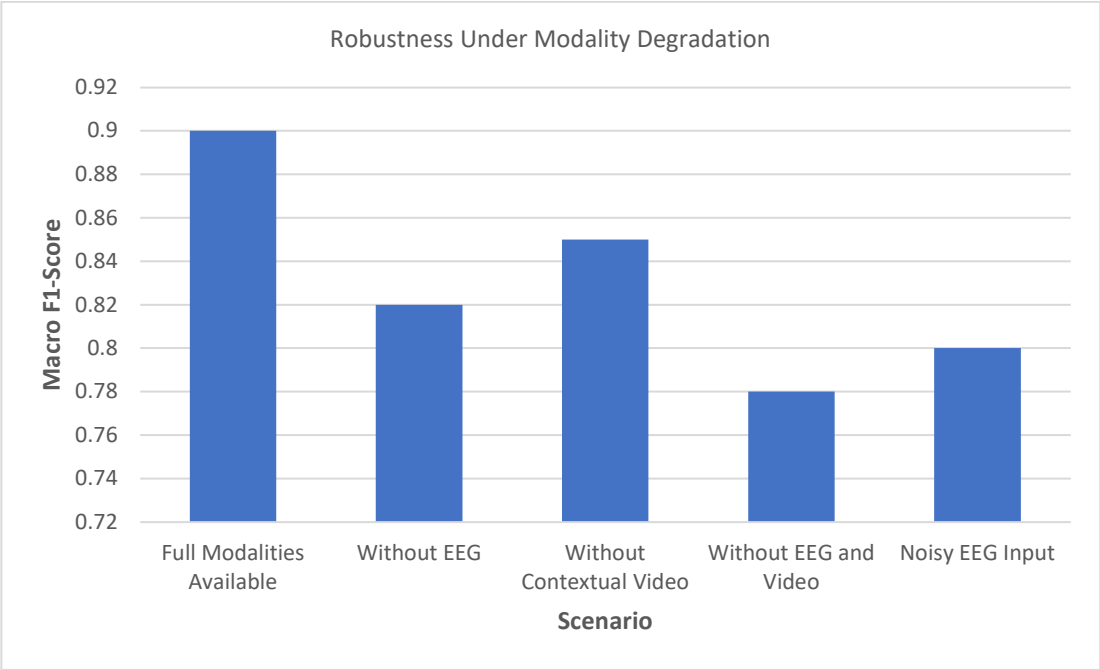


Figure 9: Macro-averaged F1-score under modality degradation scenarios

The most pronounced performance drop occurs when EEG data are removed or heavily corrupted, reflecting the importance of neurophysiological information for detecting subtle concussion-related disruptions. Removal of contextual video leads to a smaller but noticeable decline, indicating that visual cues provide complementary evidence rather than serving as a primary predictor. Even under combined modality loss, the framework maintains reasonable discrimination capability, underscoring its suitability for deployment in dynamic and imperfect sensing environments. Overall, the results presented in Table 12 and Figure 8 confirm that the proposed framework maintains robust performance under realistic operational constraints while preserving reliable concussion risk stratification

4.2 Qualitative Results

4.2.1 Explainability and Feature Attribution Analysis

Explainability analysis of the predictions for concussion risk transparency and clinical meaningfulness is discussed, using SHAP-based feature attribution and Transformer attention visualization. Explanations are provided at the modality and feature levels for individual impact events to explain how the modalities contributed to predicted risk category. Attention maps from the multimodal fusion module depict how modality prioritization varies during the inference. Unique and clinically relevant attribution profiles are evident across the levels of concussion risk. Rotational acceleration

peaks, asymmetrically altered EEG frequencies, and unfavorable collision angles on video images were all strong indicators of high concussion risk, while cumulative biomechanical exposure with sustained, subtle neurophysiological changes correlated with moderate risk. Symmetric, low-magnitude kinematics, minimal changes in EEG activity, and poor weighting of contextual/visual information were present in low-risk situations. An overall profile of dominant contributing features for each risk level is summarized in Table 14.

Table 14. Summary of Explainability Findings Across Concussion Risk Levels

Predicted Risk Level	Dominant Modalities	Attributed Features	Clinical Interpretation
High risk	Kinematics, EEG, Video	Rotational acceleration peaks, asymmetric EEG band disruptions, adverse collision angles	Strong indicators of acute biomechanical and neurophysiological injury
Moderate risk	Kinematics, EEG	Cumulative impact exposure, subtle EEG deviations	Emerging or compounded injury risk
Low risk	Kinematics, EEG	Symmetric low-magnitude motion, stable EEG responses	Consistent with non-injurious or benign events

Besides a qualitative evaluation, the robustness of explanation is also assessed quantitatively through SHAP Attribution Stability. It calculates how stable the feature attributions are across the k-fold cross validation. The stability is defined as the average Spearman rank correlation of SHAP value vectors of the same impact event across various folds. The results show consistently high stability scores for all risk categories, especially high stability in the high-risk prediction group which reveals that the resulting explanation is stable. Table 15 presents the overall quantitative explainability measures.

Table 15. Quantitative Explainability Results Based on SHAP Attribution Stability

Predicted Risk Level	SHAP Attribution Stability (Mean ± SD)	Interpretation
High risk	0.82 ± 0.05	Highly consistent and reliable explanations
Moderate risk	0.76 ± 0.07	Stable explanations with moderate variability
Low risk	0.79 ± 0.06	Consistent attribution of benign patterns

Collectively, qualitative and quantitative results indicate that the developed system generates a transparent and robust explanation. The identified attribution patterns conform well to known concussion biomechanics and neurophysiology. All these results contribute to the clinical validation and reliability of the explainable hybrid deep learning system.

4.2.2 Modality – Level Attention Pattern

Modality-level attention analysis examines how the multimodal Transformer dynamically allocates importance across data sources during concussion risk prediction. Attention heatmaps reveal event-dependent weighting patterns that vary across attention layers and impact conditions. Representative attention visualizations are shown in Figure 10.

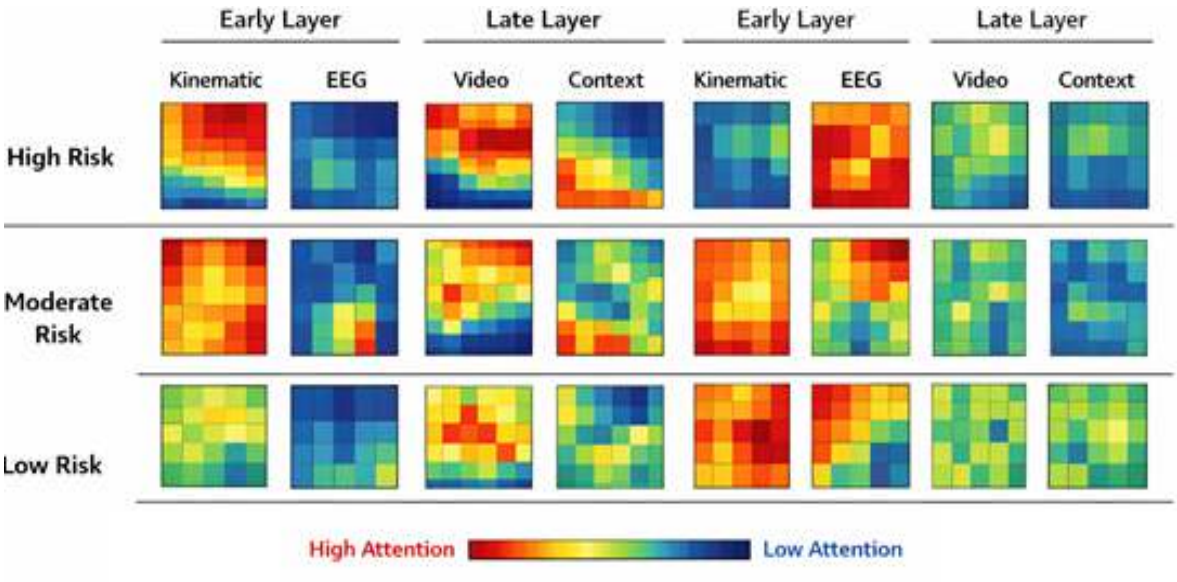


Figure 10: Modality-Level Attention Heatmaps Across Concussion Risk Levels

For high-impact events, attention weighting on kinematic and video modalities are strong in earlier fused layers, corresponding to the immediate biomechanical and visual signatures of impact event. EEG modality weights are growing at later layers to capture delayed neurophysiological signature at the beginning of the concussion. For low-risk events, the attention on different modalities is relatively balanced and symmetric, indicating healthy patterns. This selective attention also supports the context-dependent reasoning instead of invariant feature selection. Attention mechanisms capture invariants in the multimodal sequences and amplify asymmetries with injury mechanisms. The trend of modality attention by different risk groups is shown in Table 16. This validates the designed framework ability to simulate real injury mechanisms using symmetric and multi-scale multimodal fusion.

Table 16: Summary of Modality-Level Attention Patterns Across Risk Categories

Risk Category	Early-Layer Dominant Modalities	Late-Layer Dominant Modalities	Interpretation
High risk	Kinematics, Video	EEG	Immediate biomechanical impact followed by neurophysiological response
Moderate risk	Kinematics	Kinematics, EEG	Accumulated exposure with emerging neural effects
Low risk	Balanced across modalities	Balanced across modalities	Symmetric, stable, non-injurious patterns

4.2.3 Case – Based Clinical Interpretation

For case-based clinical interpretation, this paper assesses how the proposed framework accommodates real-world decisions through typical impact scenario illustrations. In cases of severe impact, the system consistently produces high concussion risk probability with accurate attribution to the biomechanical severity indices like angular acceleration peak, and neural abnormality revealed in the EEG pattern. Such prediction results correspond with high confidence and low estimation of uncertainty so that critical events could be swiftly discovered by a clinician. Cases with less severity are presented with moderate risk probability and higher prediction uncertainty. Higher uncertainty value can be perceived as an alert to caution a clinician that the automated suggestion should not be trusted alone and further clinical evaluations, observation or follow-up is required. Such operation is compliant with the Safe and Accountable AI approach for medical risk decision support in which awareness of uncertainty is central. Expert-in-the-loop evaluation further asserts the clinical utility of the constructed explanation. Domain experts find the feature attributions and attention patterns

congruent with the established concussion assessment procedures and generate a set of useful pieces of evidence that assist, but not override the clinical judgment. Overall, these results evidence that the proposed explainable hybrid framework improves the trust, interpretability and safety of concussion risk prediction.

5. Discussions

The experiment results illustrate that the presented framework with symmetry-aware multiscale explainable hybrid improves the multimodal concussion-risk prediction than the unimodal and the multimodal ones. The overall prediction accuracy reaches 0.91 ± 0.02, while the AUC reaches 0.95 0.01. These high prediction andAUCvalues prove the excellent discriminating ability and reliable probabilistic separation between categories. For high-risk events detection, recall of 0.89 ± 0.03 shows a strong capacity to identify serious situations and lessen the rate of undetected acute conditions, and F1-score of 0.87 ± 0.03 for moderate risk events reveals better sensitivity of subtle neurophysiological and biomechanical features in case of typical undetected signs. The advantages of adaptive Transformer-based multimodal fusion is further evident in comparison with the two unimodal/multimodal systems using standard fused models. Multimodal CNN models with early fusion attain 0.84 prediction accuracy and 0.81 macro F1-score, and multimodal late fusion models reach 0.87 accuracy and 0.85 macro F1-score. However, our method achieves 0.91 prediction accuracy and 0.90 macro F1-score, which better balance precision and recall across all risk classes. Ablation analysis confirmed the effectiveness of multimodality. Removing EEG feature leads to a drop in macro F1-score from 0.90 to 0.82. Using direct feature concatenation instead of Transformer-based fusion leads to a decrease of prediction accuracy from 0.91 to 0.83. This suggests that neurophysiological features are of critical importance, and the adaptive attention-based fusion mechanism is indeed powerful.

Robustness analysis indicated relatively constant participant-independent performance with little variability in cross-validation (0.02 for both accuracy and macro F1 score). Even under missing modality, performance decayed gracefully, showing robust performance to partial sensor failure, where macro F1 score was approximately 0.78 even when both EEG and video data inputs were removed. Explainability analysis illustrated steady attribution consistent with SHAP attribution stability (0.82 for high-risk predictions), indicating consistency in model reasoning to biomechanic and neurophysiological features critical for clinical insight. The following several constraints are of key importance. External validation through multi-center independent cohorts was not available in this study and all evaluation were done through within-cohort subject-independent cross-validation. Potential sampling biases prevented adequate representation of various demographic characteristics and limitations in acquisition environment may reduce robustness. Performance of wearable sensors can be variable across various health care systems, patient groups, wearable sensor types, and institutional data collection processes

Additional challenges exist in practical deployment. Variability in calibration among wearable sensors, synchronization issues across modalities, environment noise, motion artifacts and patient-specific physiological variation may lead to a decrease in predictive stability when moving out of the controlled test environment. The computational burden, memory consumption, and inference delay are higher compared to simpler fusion methods due to transformer-based multimodal fusion, and these may prohibit real-time monitoring and resource-constrained devices. Additionally, explainability requirements, clinician trust, integration into clinician workflow and regulatory approval may be barriers for clinical adoption. Nevertheless, the framework showed predictive performance, modality robustness and interpretable reasoning that warrants further development and consideration for future clinical decision support applications.

6. Conclusion and Future Work

We developed a symmetry-aware multimodal explainable hybrid framework using multimodal data fusion to predict concussion risk. The framework combines kinematic signals, EEG recordings, contextual metadata, and video stream through attention with Transformer based network for prediction. The experiments achieve significant prediction results (accuracy of 0.91, macro F1-score of 0.90, and AUC of 0.95), with stability across subject-independent test folds. The results demonstrate strong ability to discriminate against the severity level of concussion risk. In terms of explainability, we use SHAP based attribution and attention visualization to interpret how multimodal features contribute to the prediction. It turns out that the attributed multimodal features correlate well with known biomechanical and neurophysiological mechanisms of concussion, thus suggesting the interpretability and potential clinical relevance. Note that the results achieved were from retrospective assessment and should be considered within the context of collected data. Many shortcomings still exist: The size of data, the imbalance of classes, the lack of diversity in cohort make it

difficult to guarantee the generalization to unseen individuals beyond current population. Evaluation relies on cross-validation rather than external independent cohorts, and performance under real-world conditions may be influenced by sensor noise, missing modalities, and variability in acquisition environments. Computational complexity associated with Transformer-based fusion also introduces challenges for real-time and resource-constrained deployment. Future work will focus on validation using larger and more diverse datasets, including longitudinal data to capture cumulative concussion effects. External benchmarking and prospective studies involving expert-in-the-loop evaluation are necessary to establish clinical reliability and practical utility. Model optimization for efficient inference will support deployment on wearable and edge devices, while exploration of additional modalities such as eye-tracking or speech signals may further enhance predictive capability. These directions aim to strengthen robustness, scalability, and real-world applicability while ensuring that conclusions remain aligned with the current evidence.

Author Contributions: Conceptualization, A.N.B., A.B. and R.O.O.; methodology, A.N.B., A.B. and R.O.O.; software, D.P.O., F.A.A. and J.B.A.; validation, A.F.K., S.B.M and D.P; formal analysis, A.N.B., A.B. and R.O.O.; investigation, A.N.B, R.O.O. and J.B.A; resources, A.N.B., A.B., R.O.O., D.P.O, A.F.K., S.B.M., D.P., R.W.B., and J.B.A.; data curation, A.N.B, A.B., R.O.O., J.B.A. F.A.A.; writing—original draft preparation, A.N.B. and R.O.O.; writing review and editing, A.B., D.P.O., R.W.B. and J.B.A..; visualization, R.O.O., F.A.A and D.P.; supervision, A.N.B, R.O.O., R.W.B.. and J.B.A.; project administration, A.N.B. and R.O.O. All authors have read and agreed to the published version of the manuscript

Competing interests: The authors declare no competing interests.

Funding Declaration: The research received no funding from any public, commercial, or not-for-profit funding agency.

Ethics approval: Not Applicable

Data Availability: The datasets used in this study are publicly available through established online repositories and were accessed between 5 December 2025 and 16 December 2025. Head-impact kinematic data were obtained from the NCAA–DoD CARE Consortium via the FITBIR repository (https://fitbir.nih.gov). EEG data were accessed from public repositories hosted on PhysioNet (https://physionet.org). Video data were sourced from publicly available sports collision and action datasets, including the UCF Sports Action Dataset (https://www.crcv.ucf.edu/data/UCF_Sports_Action.php). Contextual and event-level metadata were derived from the annotations accompanying these datasets. The integrated multimodal dataset used in this study was created through preprocessing, temporal alignment, and synchronization of these public data sources. All datasets are accessible online through their respective repositories.

7. References

Abrol, A., Fu, Z., Du, Y., & Calhoun, V. D. (2019). Multimodal data fusion of deep learning and dynamic functional connectivity features to predict Alzheimer's disease progression. *Proceedings of the 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 4409–4413. https://doi.org/10.1109/EMBC.2019.8856500

AlNuaimi, H., Mohamed Umar, M., Memon, Q. A., & Al Ahmad, M. (2024). Advanced impact measurement method of the head using piezoelectric sensors. *IEEE Sensors Journal, 24*(15), 24570–24580. https://doi.org/10.1109/JSEN.2024.3409877

Asani, F., Patel, B., Srikanthan, S., & Agu, E. (2023). BioscoreNet: Traumatic brain injury detection using a multimodal self-attention fusion neural network and a passive bioscore monitoring framework from smartphone sensor data. *Smart Health, 27*, 100352. https://doi.org/10.1016/j.smhl.2022.100352

Babatunde, A. N., Balogun, B. F., Ajagbe, S. A., Akpan, E. E., Ogundokun, R. O., Ogie, P. I., Isiaka, S. O., & Mudali, P. (2025a). Breast cancer classification using DenseNet121 and k-means segmentation with augmented data. *Informatica, 49*(27). https://doi.org/10.31449/inf.v49i27.8332

Babatunde, A. N., Balogun, B. F., Babatunde, A. J., Isiaka, S. O., Awotunde, J. B., & Adeniyi, A. E. (2025b). Improving skin lesion classification using k-means, transfer learning, and MobileNet architecture. *Network Modeling Analysis in Health Informatics and Bioinformatics, 14*(1), 70. https://doi.org/10.1007/s13721-025-00569-3

De Luca, D., Moccia, S., Lupori, L., Mazziotti, R., Pizzorusso, T., & Micera, S. (2023). Convolutional neural network classifies visual stimuli from cortical response recorded with wide-field imaging in mice. *Journal of Neural Engineering, 20*(2), 026031. https://doi.org/10.1088/1741-2552/acc2e7

Dell, K. C., Grossner, E. C., Staph, J., Schatz, P., & Hillary, F. G. (2021). A population-based study of pre-existing health conditions in traumatic brain injury. *Neurotrauma Reports, 2*(1), 255–269. https://doi.org/10.1089/neur.2020.0065

Du, W., Wang, S., Chi, H., Chen, X., Li, W., Wan, H., Hou, D., & Cao, Y. (2024). Evaluating head impact intensities and accelerations using flexible wearable sensors for traumatic brain injury assessment. *Sensors and Actuators A: Physical, 373*, 115443. https://doi.org/10.1016/j.sna.2024.115443

Edelstein, R., Gutterman, S., Newman, B., & Van Horn, J. D. (2024). Assessment of sports concussion in female athletes: A role for neuroinformatics? *Neuroinformatics, 22*(4), 607–618. https://doi.org/10.1007/s12021-024-09680-8

Ghonge, M., Shukla, V. K., Pradeep, N., & Solanki, R. K. (2025). Recent advances in multimodal deep learning for stress prediction: Toward cycle-aware and gender-sensitive health analytics. *EPJ Web of Conferences, 341*, 01059. https://doi.org/10.1051/epjconf/202534101059

Iparraguirre-Villanueva, O., Gonzales-Huaman, J., Machuca-Solano, J., & Ruiz-Alvarado, J. (2024). Improving industrial security device detection with convolutional neural networks. *Indonesian Journal of Electrical Engineering and Computer Science, 34*(3), 1935–1943. https://doi.org/10.11591/ijeecs.v34.i3.pp1935-1943

Isiaka, S. O., Babatunde, R. S., & Isiaka, R. M. (2024). Exploring artificial intelligence technologies in predictive medicine: A systematic review. *KASU Journal of Computer Science, 1*(2), 366–377. https://doi.org/10.47514/kjcs/2024.1.2.0014

Jolly, A. E., Bălăeţ, M., Azor, A., Friedland, D., Sandrone, S., Graham, N. S. N., Zimmerman, K., & Sharp, D. J. (2021). Detecting axonal injury in individual patients after traumatic brain injury. *Brain, 144*(1), 92–113. https://doi.org/10.1093/brain/awaa372

Koochaki, F., Shamsi, F., & Najafizadeh, L. (2020). Detecting mild traumatic brain injury by learning spatiotemporal characteristics of widefield calcium imaging data using deep learning. *Proceedings of the 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 2917–2920. https://doi.org/10.1109/EMBC44109.2020.9175327

Liu, J., Capurro, D., Nguyen, A., & Verspoor, K. (2023). Attention-based multimodal fusion with contrast for robust clinical prediction in the face of missing modalities. *Journal of Biomedical Informatics, 145*, 104466. https://doi.org/10.1016/j.jbi.2023.104466

Mahajan, C., Sengupta, D., Kapoor, I., Prabhakar, H., Kumar, V., Purohit, S., Priya, V., Srivastava, S., Thakur, D., Karnik, H., Sati, H. C., & Kalaivani, M. (2023). Evaluation of the GCS–Pupils score for prognosis in traumatic brain injury: The COMA study. *Brain Injury, 37*(9), 1041–1047. https://doi.org/10.1080/02699052.2023.2227943

Mahootiha, M., Tak, D., Ye, Z., Zapaishchykova, A., Likitlersuang, J., Climent Pardo, J. C., Boyd, A., Vajapeyam, S., Chopra, R., Prabhu, S. P., Liu, K. X., Elhalawani, H., Nabavizadeh, A., Familiar, A., Mueller, S., Aerts, H. J. W. L., Bandopadhayay, P., Ligon, K. L., Haas-Kogan, D., … Kann, B. H. (2025). Multimodal deep learning improves recurrence risk prediction in pediatric low-grade gliomas. *Neuro-Oncology, 27*(1), 277–290. https://doi.org/10.1093/neuonc/noae173

Mazumder, B., Tripathy, D. K., Yeates, K. O., Beauchamp, M. H., Craig, W., Doan, Q., Freedman, S. B., Lebel, C., Zemek, R., Ware, A. L., & Hye Ye, D. (2023). Multimodal deep learning for pediatric mild traumatic brain injury detection. *Proceedings of the IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, 1–4. https://doi.org/10.1109/BHI58575.2023.10313520

McLeod, J., Thanjavur, K., Sattari, S., Babul, A., Hristopulos, D. T., & Virji-Babul, N. (2025). Linking a deep learning model for concussion classification with reorganization of large-scale brain networks in female youth. *Bioengineering, 12*(9), 986. https://doi.org/10.3390/bioengineering12090986

Menghani, R. R., Das, A., & Kraft, R. H. (2023). A sensor-enabled cloud-based computing platform for computational brain biomechanics. *Computer Methods and Programs in Biomedicine, 233*, 107470. https://doi.org/10.1016/j.cmpb.2023.107470

Nayak, C., Patel, S., Mahajan, A., & Rathore, M. (2025). Optimizing mental health diagnostics with hybrid deep learning and multimodal data fusion. *International Journal of Environmental Sciences, 11*(5s), 13–27. https://doi.org/10.64252/sfqnv852

Nishizawa, T., Maldjian, T., Jiao, Z., & Duong, T. Q. (2025). Attention-based multimodal deep learning for interpretable and generalizable prediction of pathological complete response in breast cancer. *Journal of Translational Medicine, 23*(1), 774. https://doi.org/10.1186/s12967-025-06617-w

Salsabilian, S., Bibineyshvili, Y., Margolis, D. J., & Najafizadeh, L. (2022). Identifying mild traumatic brain injury using measures of frequency-specified networks. *Journal of Neural Engineering, 19*(5), 056033. https://doi.org/10.1088/1741-2552/ac954e

Salsabilian, S., & Najafizadeh, L. (2021). An adversarial variational autoencoder approach toward transfer learning for mild traumatic brain injury identification. *Proceedings of the 10th International IEEE/EMBS Conference on Neural Engineering (NER)*, 408–411. https://doi.org/10.1109/NER49283.2021.9441372

Shajari, S., Kuruvinashetti, K., Komeili, A., & Sundararaj, U. (2023). The emergence of AI-based wearable sensors for digital health technology: A review. *Sensors, 23*(23), 9498. https://doi.org/10.3390/s23239498

Sprint, G., Cook, D. J., Schmitter-Edgecombe, M., & Holder, L. B. (2022). Multimodal fusion of smart home and text-based behavior markers for clinical assessment prediction. *ACM Transactions on Computing for Healthcare, 3*(4), 1–25. https://doi.org/10.1145/3531231

Teoh, J. R., Dong, J., Zuo, X., Lai, K. W., Hasikin, K., & Wu, X. (2024). Advancing healthcare through multimodal data fusion: A comprehensive review of techniques and applications. *PeerJ Computer Science, 10*, e2298. https://doi.org/10.7717/peerj-cs.2298

Tran, A. T., Zeevi, T., Haider, S. P., Abou Karam, G., Berson, E. R., Tharmaseelan, H., Qureshi, A. I., Sanelli, P. C., Werring, D. J., Malhotra, A., Petersen, N. H., De Havenon, A., Falcone, G. J., Sheth, K. N., & Payabvash, S. (2024). Uncertainty-aware deep-learning model for prediction of supratentorial hematoma expansion from admission noncontrast head computed tomography scan. *NPJ Digital Medicine, 7*(1), 26. https://doi.org/10.1038/s41746-024-01007-w

Vylala, A., Plakkottu Radhakrishnan, B., & Balakrishnan Kadan, A. (2025). Early prediction and risk analysis using hybrid deep learning techniques in multimodal biomedical imaging. *Developmental Neurobiology, 85*(4), e23001. https://doi.org/10.1002/dneu.23001

Ye, J., Hai, J., Song, J., & Wang, Z. (2023). Multimodal data hybrid fusion and natural language processing for clinical prediction models. *Health Informatics*. https://doi.org/10.1101/2023.08.24.23294597

Zhan, X., Oeur, A., Liu, Y., Zeineh, M. M., Grant, G. A., Margulies, S. S., & Camarillo, D. B. (2022). Translational models of mild traumatic brain injury tissue biomechanics. *Current Opinion in Biomedical Engineering, 24*, 100422. https://doi.org/10.1016/j.cobme.2022.100422

Zhuang, Y., Yang, Q., Han, T., O'Malley, R., Kumar, A., Gerald, R. E., & Huang, J. (2021). Fiber optic sensor embedded smart helmet for real-time impact sensing and analysis through machine learning. *Journal of Neuroscience Methods, 351*, 109073. https://doi.org/10.1016/j.jneumeth.2021.109073

Zompanti, A., Basoli, F., Saggio, G., Mattioli, F., Sabatini, A., Grasso, S., Marino, M., Longo, U. G., Trombetta, M., & Santonico, M. (2024). Design, calibration, and morphological characterization of a flexible sensor with adjustable chemical sensitivity and possible applications to sports medicine. *Sensors, 24*(19), 6182. https://doi.org/10.3390/s24196182