Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

ML-Assisted Cognitive Characterization of Human Problem Solving Behaviors through Visuo-Spatial Game

Dipshikha Podder
Indian Institute of Technology Kharagpur
Sonali Aatrai
Indian Institute of Technology Kharagpur
Aritra Hazra
Indian Institute of Technology Kharagpur
Rajlakshmi Guha
rajg@cet.iitkgp.ac.in

Indian Institute of Technology Kharagpur
Partha Pratim Chakrabarti
Indian Institute of Technology Kharagpur

Article

Keywords: Visuo-Spatial Games, Cognitive Problem Solving, Machine Learning, Eye Parameters, Eye Tracking

Posted Date: September 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8794593/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

ML-Assisted Cognitive Characterization of Human Problem Solving Behaviors through Visuo-Spatial Game

Dipshikha Podder[1], Sonali Aatrai[2], Aritra Hazra[3], Rajlakshmi Guha(Corresponding author)[1*], Partha Pratim Chakrabarti[3]

[1*]Rekhi Centre of Excellence for the Science of Happiness, Indian Institute of Technology Kharagpur, Kharagpur, 721302, West Bengal, India.

[2]Advanced Technology Development Centre, Indian Institute of Technology Kharagpur, Kharagpur, 721302, West Bengal, India.

[3]Computer Science and Engineering, Indian Institute of Technology Kharagpur, Kharagpur, 721302, West Bengal, India.

*Corresponding author(s). E-mail(s): rajg@cet.iitkgp.ac.in;

Abstract

This study aimed to categorize the players in a visuo-spatial task based on their gaze behavior and gameplay behavior. This article used the ensemble-based Random Forest classifier and regressor on an available eye-tracker based dataset. We used the gameplay behavior to generate the fixation and saccade data. The findings from this study suggested that the 'Solvers' who solved the puzzle and the 'Non-solvers' who quit the game had a clear distinction in their gaze behavior, which was revealed from the results of the classifier. The results from both the classifier and the regressor suggested that the Solvers can be further categorized into two subcategories, the 'Experts', whose gameplay was strategic and the 'Persistent players', whose gameplay was erratic. Similarly, the Non-solvers can also be categorized into two types, the 'Triers' who intend to solve the problem but, for lack of understanding, showed exploratory behavior for a longer period and finally quit the game, and the 'Quitter' who had no intention to solve the puzzle and did not shift their approach efficiently.

Keywords: Visuo-Spatial Games, Cognitive Problem Solving, Machine Learning, Eye Parameters, Eye Tracking

1 Introduction

Problem solving, in general, involves a cognitively demanding thinking capability and often utilizes trial-and-error based methods, leveraging a well-defined algorithmic or heuristic approach to reach its solution. It also employs manifold tactics such as seeking, redefining, reforming or condensing to produce a solution (Sinaga et al. 2023) – thereby surfacing the human reasoning and cognitive learning aspects in a broader dimension. A large amount of insights (Robertson 2016) and analogies (Reed 1992) may be drawn about the human cognitive processes through visual behavior during problem solving. Accordingly, the ability of humans to solve problems and apply different solution strategies has been tested on a variety of visuo-spatial problems (McCloskey 1985; MacGregor et al. 1999; Landgraf et al. 2011). For example, games/puzzles like Jigsaw(Kocijan et al. 2017) or 8-tiles (Eivazi and Bednarik 2010; Khakifirooz 2015) primarily try to assess the spatial reasoning skill, planning and strategy.

Recent psychological analyzes of human problem solving vis-a-vis syntheses of computer-generated intelligent problem solving both involve viewing problem solving as a form of information processing by understanding the configurations and navigating through them forming solution pathways (Graham 2025). On one hand, studying the eye movements to comprehend cognitive processes involved in problem solving by humans has been gaining traction in the past two decades (Zang et al. 2022; Tang et al. 2016; Guo et al. 2024). On the other hand, artificial intelligence (AI) or machine learning (ML) models are increasingly being deployed to understand the intelligent, experience/data-driven learned approaches for completing similar tasks.

Though there have been multiple studies on understanding human problem solving and machine problem solving in isolation (Liu et al. 2017; Eivazi and Bednarik 2010; Bengio et al. 2021; Mazyavkina et al. 2021), there seems to be a lack of effort in providing a comparative landscape between human and artificial mechanisms for solving such visuo-spatial tasks. Such a gap may also be attributed to the human information processing and long-short term memory features. Human problem-solving is manifestly fallible as they are prone to many errors, false starts, apparent inconsistencies (manifested upon hitting dead ends), and failures to progress (manifested upon getting stuck). However, the identification and prediction of human stages of progression inside the visuo-spatial problem-solving process is challenging and may vary largely across individuals (Atkinson and Nardini 2008).

The available studies used two assessment mechanisms to capture the problem-solving behavior: (a) the direct method, where the behavior was captured through some biomarkers or direct process-level gameplay data and (b) the indirect method, where the behavior was captured through any standard questionnaire or the think-aloud protocol. In the indirect method, the think-aloud protocol is used while solving the problem, or the experimenters presented the questionnaires in front of the participants after they finished the problem. Problem Solving Inventory (PSI) (Ariff et al. 2017), Programme for International Assessment of Adult Competencies (PIAAC)[1] (Maddox et al. 2018), or Programme for International Students'

[1] https://www.oecd.org/en/about/programmes/piaac.html

Assessment (PISA)[2] (Eichmann et al. 2020; Pejić and Molcer 2019; Le Hebel et al. 2016) are the example of such questionnaires.

To capture the behavior directly, different types of bio markers can be used, such as – (a) Electroencephalography(EEG) as demonstrated by Bastiaansen and Hagoort (Bastiaansen and Hagoort 2006) who captured the change in theta and alpha power in brain while solving a problem; (b) Functional Magnetic Resonance Imaging (fMRI) which Goel and Dolan (Goel and Dolan 2004) used to differentiate the involvement of left prefrontal cortex in inductive and deductive reasoning; (c) Eye tracker as explained by Eivazi and Bednarik (Eivazi and Bednarik 2010) by capturing the problem solving behavior and performance. Besides the biomarkers, to assess the game-playing strategy of an individual, some studies used different types of games or puzzles, such as the Cosmiclean (Udeozor et al. 2021), Tower of Hanoi (Resing et al. 2020), Jigsaw Puzzles (McInnes et al. 2009), Number Puzzles(Huang et al. 2007), Circuit Puzzles (Wang et al. 2021), and 8-tiles Puzzle (Eivazi and Bednarik 2010; Khakifirooz 2015), Sudoku (Kalia et al. 2019). Though the mentioned works used game behavior of the participants, except the work that used Cosmiclean (Udeozor et al. 2021), others did not find the behavioral difference between the solvers and 'Non-solvers'. Udeozora et. al. (Udeozor et al. 2021) also found the difference between two groups of solvers – one who solved more levels than others who could not progress beyond a few levels.

Though questionnaires are capable of giving some kind of insight into the problem-solving behavior, it is subjective to the participant. Nisbett and Wilson (Nisbett and Wilson 1977) explained that participants are sometimes unaware of the stimulus and the cognition behind their response, and their response depends on theories and judgments about the stimulus. Moreover, problem-solving is a dynamic process where the strategy changes from time to time; biomarkers can effectively capture these changes as they record the real-time data. Many problem-solving strategies, like insights, are implicit, which can only be recorded through direct measures. Insight is known as the sudden realization of the correct solution. Thomas and Lleras (Thomas and Lleras 2009) showed that gaze shift can capture when the insight appears.

While EEG and fMRI capture the brain mechanisms and cognitive load, eye tracking records the in-process behavioral details, including where the participants look and how long they look at a particular information. A longer fixation typically indicates deeper cognitive processing, whereas the number of fixations reflects the amount of information being processed. The length of saccades suggests whether participants are engaging in a global search or focusing locally, while revisits indicate re-evaluation or uncertainty. Overall, eye tracking provides rich and process-level insights into the problem-solving behavior.

This article enables a novel yet different approach to gain insights into the human problem-solving process by employing an equivalent capable computational model guided by parametric eye-markers. Literature showed that the eye tracker can be effectively used to capture cognitive behavior, hence in this study, we used an available eye tracker based dataset (Aatrai et al. 2023). In the pre-processing part, we incorporated the game-play behavior by using eight different sliding windows to generate

[2]https://www.oecd.org/en/about/programmes/pisa.html

the time series gaze dataset. We investigate (both reason and predict) human behavior while solving visuo-spatial tasks/puzzles by leveraging artificially learned models. The primary objectives of this work are mentioned below.

- Classifying the *'Solver'* and the *'Non-solver'* groups as captured from the game-play behavior as well as further categorizing 'Solvers' into *'Expert'* and *'Persistent'* groups, and 'Non-solvers' into *'Triers'* and *'Quitters'* groups.
- Detecting the earliest *saddle-point* in game-play from where 'Solver' vs. 'Non-solver' identification stabilizes and thereby formulating the segregation of group-wise dynamics with respect to the saddle-point.
- Predicting the remaining game-play behavior, like the ability of reaching to a solution and the number of steps to finish/quit, by the adoption of predictive models.
- Identify the behavioral differences between the 'Solver' and 'Non-solvers' and their subcategories.

This study used ensemble-based Random-Forest classifier and regressor in classifying the groups and predicting the remaining gameplay. To predict the remaining gameplay, the remaining number of moves was predicted using Random-Forest regressor. This study first divided the dataset into four groups based on the gameplay behavior and solvability behavior, and then, through analyzing the results obtained with the classifier and regressor, we established the existence of four groups in the dataset. The results were analyzed in two ways, i.e., individual analysis and group-wise analysis. For the individual analysis, the predicted classification accuracy and predicted remaining moves were plotted over gameplay. For the group analysis, Mean Absolute Error (MAE) and Standard deviation were calculated. The result presented that the dataset had the gaze data of four types of participants, a group that solved very fast and their gameplay strategy settled early, a group that required time to settle in their game play behavior and finally solved the puzzle, a group who could not solve the puzzle and another group who left the game but developed some insight into the game. A graphical representation of the methodology used in this study is given in Figure 1.

The rest of the article is organized as follows. Section 2 introduces the required background with the details of the game used for our study, eye features and cognitive modeling of human problem-solving behavior and introduces the basics of machine learning models used in our study. The complete methodology and implementation setup, including data preparation and modeling are detailed in Section 3. In Section 4, we establish through our experiments various classification, prediction and cognition behaviors and discuss their ramification across groups and individuals. We further discuss interesting insights obtained through this study, which relate to human cognitive problem-solving behavior in Section 4.5. Finally, Section 5 concludes the work presented in this article.

2 Background and Preliminaries

In this section, we present the necessary concepts that are required to understand cognitive problem solving including human behavioral models involving eye parameters

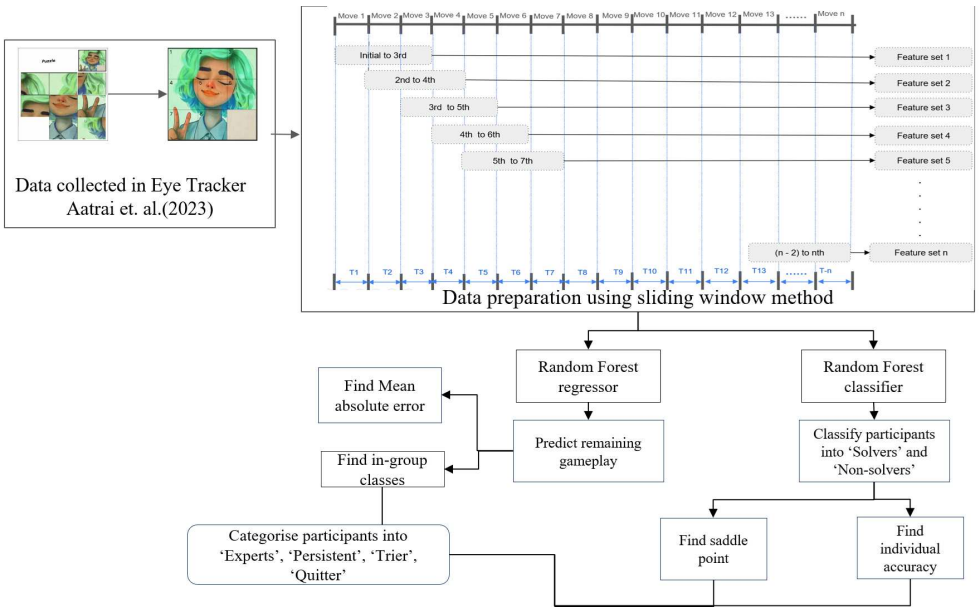


Fig. 1: Summary of our Overall Approach

as the analytic tool. Later, we also provide a preliminary outline of machine learning models that we had used in our experiments for further cognitive predictions.

2.1 Cognitive Models in Problem Solving

The ability of humans to comprehend, reason and solve problems has been studied over several decades. A large amount of insight can be drawn about the human cognitive processes through visual behavior, as well as the application of a variety of solution strategies during problem-solving. Often, learning aids the problem-solving behavior as solving a particular problem may provide a basis for generalization to other problems, as well as the experience gathered specific to that problem. Human working memory (both long-term and short-term) and its utilization and retrieval also play a pivotal role in such cognitive problem-solving. Based on all these aspects, different problem-solving models have been introduced in the past as described below.

Poyla's Model

According to this model, introduced by Poyla in 1957 (Polya 1957, 2014), the cognitive behaviors involved in problem solving are – (a) identifying the problem, (b) planning the solution, (c) implementing the solution, and finally (d) checking the solution for reasonableness and consistency. Identification of the problem requires comprehension of the problem, including an understanding of the initial and the goal configurations. Planning the solution may involve recall from the memory (in case of navigating through known solution pathways) as well as involving the working memory to build strategies. Human approaches to problem solving can either follow a

heuristic approach, i.e., focusing on solving the local parts of the problem, or take a global (holistic) approach, i.e., considering the whole problem at once to get a solution, or a combination of both using decomposition of the overall problem into parts and recomposition of the partly solved instances.

Newell-Simon's Model

This is the first model for *cognitive characterization* of human problem solving, introduced by Newell and Simon in 1972 (Newell et al. 1972), states that people take small steps to solve a problem while thinking about the overall problem. It is also known as the general problem-solving model or the information processing model, which states that human solves problems by searching through their states/configurations in the problem space and uses means-end analysis, heuristics, and walking backwards as strategies to navigate through various pathways in order to reach the goal state from the initial state. The choice of paths and diversions for a human learner explicitly marks the progress being made, including reaching blocks and dead ends and then backtracking for searching alternative solutions as part of the process of state-space search. These markers may be witnessed as a signature of the learning pathway and act as a guide to the problem-solving behavior. The nature of state-space traversal for problem-solving shall provide essential cues to the problem-solving process.

Gestalt's Model

This model was studied since 1912 (Koffka 1912) when most of the Gestalt psychologists talked about perception and some of them (like Max Wertheimer in 1945) (Wertheimer 1945) worked in the area of human problem solving and cognitive skills. It introduces the notion of developing insights to germinate conscious and sudden solution strategies from proper representations of the problem. In this context, they talked about two different ways of approaching the goal-oriented situation – (a) the reproductive thinking (primarily responding to the given problem from past experience) and (b) the productive thinking (generating newer inferences beyond previously learned associations).

Kahneman's Model

This model, introduced in various forms between 1970 to 2000 (Sloman 1996; Stanovich and West 2000; Kahneman and Frederick 2000), proposes two distinct modes of thinking: (a) The Intuitive System (System 1), which is fast, automatic, effortless, nonconscious, intuitive, and often associative; and (b) The Deliberative System (System 2), which is slow, deliberative, controlled, effortful, conscious, and logical / rule-based. System 1 relies on heuristics and is prone to biases, while System 2 engages in conscious reasoning to overcome these biases and make more deliberate decisions. Understanding the interplay between these systems helps in explaining why people make certain judgments and decisions, and how to make better choices by engaging System 2 when necessary. This model provides an idea about how our thoughts and insights may arise from the interplay of these two different systematic ways. Mostly, the players involve both processes while solving a problem.

2.2 Eye Parameters in Cognitive Problem Solving

Studying the eye movements to comprehend the cognitive processes involved in reasoning and problem solving have been gaining traction in the past two decades (Thomas and Lleras 2007; Aatrai et al. 2024). Moreover, eye-tracking is non-invasive and does not require any additional effort or training on the part of the subject. Thus, rather than being disruptive, eye movement parameters bring integral aspects of the human problem-solving process.

Cognitive Problem Solving using Eye Tracking

Körner and Gilchrist (2004) presented the role of eye movements in structuring information for solving a visuo-spatial reasoning task. Meijering, Van Rijn, Taatgen, and Verbrugge showed (Meijering et al. 2012) through eye movements that humans tend to follow a strategy similar to causal reasoning in a strategic game scenario, even though such an approach may not always be optimal. Through eye movements and a cognitive model, Nyamsuren and Taatgen (2013) showed that both top-down and bottom-up strategies were involved in parallel in a complex problem-solving task. Eye movements were used by Evans and Ball (2010) to show that reasoning affected participants' choice in the Wason selection task. Körner (2011) presented observations indicating separate search and reasoning processes in graph comprehension. Eye-tracking patterns in behavioral and problem-solving studies are usually considered on specific regions on the stimuli, called the Areas of Interest (AOIs) (eg. (Kiefer et al. 2014; Roy et al. 2017)). Choosing, delineating and measuring the reliability of AOIs can be difficult (Goldberg and Helfman 2010; Holmqvist et al. 2011).

Eye Tracking Parameters

There are two types of eye parameters captured through eye trackers, gaze-dependent and gaze-independent. Gaze-dependent parameters include fixations, saccades, and smooth pursuit, and gaze-independent parameters include pupil dilation and blink rate. Fixations are eye gazes that usually last longer than 100 milliseconds. Saccades are rapid eye movements between fixations, typically lasting $80 - 100$ milliseconds. Smooth pursuit is a type of eye movement in which the eyes track a moving object continuously. Fixation-based features such as time to first fixation, fixation count, first fixation duration, and total fixation duration, etc., have been widely used in the literature to assess problem-solving behavior (Nasreen et al. 2022, 2025; Aatrai et al. 2023, 2024). Among saccadic features, peak velocity is a very important measure; other commonly used features include saccadic amplitude, saccadic velocity, saccadic duration, and the ratio of total saccadic duration to total scanning duration. Smooth pursuit movements are closely related to attention, visual memory, and processing speed, and therefore, they are not frequently used to assess problem-solving behavior. On the other hand, pupil dilation and blink rate have been applied in problem-solving research. Though they do not directly measure problem-solving behavior, they provide valuable information about problem difficulty and cognitive load. The specific parameters used in this study are detailed in Section 3.1.

Gaze Data Extraction Process using Eye Tracker

An eye tracker typically captures gaze co-ordinates at a high sampling rate (60Hz-1000Hz). The raw data usually includes Timestamps, Gaze co-ordinates(x,y), Pupil Diameter, and Validity or Confidence Score. Before extracting the raw data, researchers clean the raw data first, where they remove the blinks, filter noise, interpolate short gaps, and finally, they extract the two types of gaze data, fixations and saccades. For gaze extraction, many gaze classification algorithms are commonly used, which are mentioned below.

- ***I-DT (Identification of Dispersion Threshold):*** Groups consecutive gaze points that stay within a small spatial area for a minimum duration.
- ***I-VT (Identification of Velocity Threshold):*** Compares point-to-point gaze velocity to a threshold (low = fixation, high = saccade)
- ***I-AVT (Identification of Angular Velocity Threshold):*** Uses angular velocity between consecutive gaze vectors to classify fixations and saccades
- ***I-HMM (Identification of Hidden Markov Model):*** Probabilistic state modeling of gaze transitions
- ***I-MST (Identification of Minimum spanning Tree):*** Uses graph clustering of gaze points to detect fixations
- ***I-AOI (Identification of Area of Interest):*** Classifies gaze points according to predefined regions on the stimulus (AOIs). Some studies consider the whole stimulus to extract the gaze features, while others consider predefined AOIs to extract the gaze features.

2.3 Visuo-Spatial Configuration Arrangement Puzzles

Puzzle solving is fundamental to scientific thinking and learning. Reconstruction of configurations using different strategies helps the individual to explore his knowledge base, involves critical thinking and problem-solving skills. Puzzles in this sense are generally solvable problem-solving entities comprising all the characteristics embedded with procedural problem-solving strategies. The use of analogy transfer in puzzle solving to solve real-life problems has been widely studied (Weng 2022).

Puzzle Solving

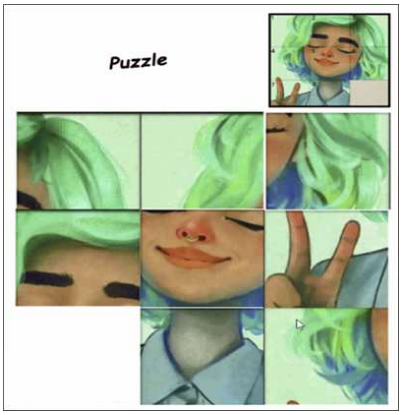
Puzzles may have several answers, and the route to the solution(s) may require the discovery of new information or the realignment and manipulation of previously known concepts into a novel configuration. The answer(s) may not be apparently obvious even when arrived at – they may be controversial or merely the most acceptable, neither right nor wrong. These may be thought of as problems, not answerable within the existing rules or frames of reference, or only answerable to a degree (not absolutely) or by the rethinking of old ideas. These problems, therefore, would seem to require all the previously discussed characteristics of creativity if they are to be solved. Puzzles make different demands upon the intellect in that they require flexibility and fluency and thus have been an important research tool to comprehend problem-solving behavior in humans.

Cognitive Explorations through Puzzle Solving

Human problem-solving primarily involves executive functions of the brain. Visuo-spatial games, such as jigsaw puzzles, demand the exploration of multiple executive functions. These include – (a) visual perception (by recognizing objects, patterns, and the orientation of lines), (b) constructional praxis (by integrating visual and motor information to assemble pieces), (c) mental rotation (by mentally rotating piece's orientation to fit them to other pieces), (d) cognitive speed and visual scanning (by sorting puzzle pieces), (e) cognitive flexibility (by switching attention between different strategies, between different puzzle pieces, and between puzzle shape, image, and color), (f) perceptual reasoning (by integrating different perceptual information to develop strategies and plans to solve the puzzle), and (g) working and episodic memory (by keeping the association between spatial location and visual patterns/images of puzzle pieces in working memory and long-term memory).

8-tile Arrangement Puzzle: *Our Case-Study*

8-tile puzzle (using a 3×3 grid) is a smaller version of the 15-tile puzzle, developed by Noyes Palmer Chapman in 1874[3], which originally uses a 4×4 grid. In an 8-tile image arrangement puzzle, an image is divided into 9 blocks, among which 1 block is removed (specifically the bottom-right tile from the actual/arranged image segment) and the rest 8 blocks are placed in the 3×3 grid randomly jumbled, keeping 1 cell vacant (white). The player needs to analyze the spatial relationship between the blocks and mentally reconstruct the actual image based on that relation. The complete image may (or may not) be shown in the top-right position outside this 3×3 grid to refer to.



A sample puzzle (with an arbitrary intermediate configuration) used in our case-study is shown above. Starting from a random initial configuration, in every move, the player is allowed only to slide one image tile to any of its left/right/top/down positions, depending on the position of the vacant space (denoted as a white block). From the edges or corners, some of such moves are obviously not allowed, particularly when the tile may slide beyond the grid area. The game is completed successfully once the player can reach the target configuration (i.e., an organized actual image)[4].

It is pertinent to note that the puzzle used in our study had a moderate level of complexity since we used 3×3 grid. Considering the face superiority effect[5] for the human face image used in the puzzle, this puzzle had a higher complexity. The puzzle

[3] https://en.wikipedia.org/wiki/15_puzzle

[4] Interestingly, the final goal configuration of the puzzle can be arranged only from half of all possible random initial configurations, and hence, it may be assumed that the players will start only from a feasible set of such initial configurations.

[5] https://en.wikipedia.org/wiki/Face_superiority_effect

used in this study had monochromatic colors, which further caused the puzzle to have increased complexity as fewer variations in colors too enhances the level of difficulty[6].

2.4 Machine Learning Techniques for Problem Solving

Recent psychological analyzes of human problem solving and syntheses of computer problem solving through artificially intelligent agents both involve viewing problem solving as a form of information processing. There seems to be a conceptual gap between how humans tackle a cognitive problem and how machine learning engines proceed to solve it. Such a gap may also be attributed to the human information processing and long short-term memory features. Human problem-solving is manifestly fallible as they are prone to many errors, false starts, apparent inconsistencies, and failures to progress ('getting stuck'). Although the nervous system is fixed, it clearly permits very flexible processing and has an important capacity for learning or self-improvement within the bounds of its fixed architecture. Hence, an equivalent capable computational model may throw adequate information in order to aid the problem-solving process of humans. Accordingly, this paves newer challenges to bridge the same because understanding the human problem-solving process will be a key component to incorporate similar cognitive intelligence factors into the automated machine learning models.

From the perspective of machine-based problem solving, two broad ways of representing problems are – (a) in state space terms and (b) in problem reduction terms. Typically, solving a problem passes through several sequences of phases starting from an initial configuration to the goal configuration. Various paths may be followed to reach the goal state from the initial state in a wide variety of ways. The choice of paths and diversions for a human learner explicitly marks the progress being made, including reaching blocks and dead-ends and then backtracking for searching alternative solutions as part of the process of state space search. These markers can be witnessed as a signature of the learning pathway and act as a guide (both in terms of predictor and estimator) to the problem-solving behavior. The nature of state space traversal for problem solving shall provide essential cues to the problem-solving process.

Since the problem solving has a strong connection with human reasoning, decision-tree based machine learning models, such as Random-Forests, etc., which provide their inherent explainable capability, seem to be most suitable to aid to such problem solving behavior. Random-Forest model is a tree-based ensemble model that relies on the bootstrap aggregation or the bagging technique (Shaik and Srinivasan 2018; Salman et al. 2024). This model trains multiple decision trees on different subsets of the dataset as well as random attribute subsets, and then combines their outputs. In the case of classification, the model collects the predictions from all the decision trees and uses a voting technique to determine the final output – the class predicted by the majority of trees becomes the final prediction of Random-Forest model. In regression problems, Random-Forest model uses averaging – it takes the outputs from all the decision trees and computes their mean to generate the final prediction. We shall use these models as generic indicators for predicting the dynamics of both individual and group behavior.

[6] https://kaayee.com/2024/05/28/how-do-you-determine-the-difficulty-level-of-a-jigsaw-puzzle

3 Details of the Proposed Methodology

This section illustrates the dataset used in our study and our proposed approaches in detail.

3.1 Eye Data used in Our Study

This study aimed to capture the process-level behavior while solving a problem; hence, it used the dataset sourced from Aatrai et al. (2023). Aatrai et al. (2023) conducted the experiment following all the regulations and guidelines given by the Institute Ethical Committee and the experiment was approved by the Institute Ethical Committee (Ethical Clearance No.: IIT/SRIC/DR/2019). The participants of the experiment were adults, hence before the experiment, informed concent were taken from the participants. That dataset contained the gameplay behavior (mouse click data) and eye gaze dataset (raw gaze data) collected while playing the 8-tile image sliding puzzle game. The dataset had the data collected from 147 participants, from where we took data of 98 participants (49 solved and 49 unsolved). The gaze features used in the current study are mentioned in Figure 2 and they were calculated using the formula given in Table A1. Aatrai et al. (2023) used a reference image to help the participants in solving the puzzle; the current study used four gaze features related to this reference image. The dataset provided us the coordinates and timestamps of the fixations, which were used along with the mouse click data to generate the final dataset that we used in the current study. All the gaze data were extracted for the time intervals between successive mouse clicks, allowing us to analyze the visual attention dynamics corresponding to each user.

3.2 Extraction of Gaze Data using Sliding Window Method

Reading, problem-solving, searching, and thinking are all dynamic behaviors. When the eye features are processed using a sliding window, it can capture how the gaze behavior is evolving over time. This study incorporated the use of a sliding window on gaze data for generating multiple, overlapping temporal observations rather than one aggregated value. In addition to a sliding window approach, the current study used the gameplay behavior. For instance, in the case of sliding window size n, first n clicks were considered, and the initial time of the first click and the final time of the nth click were checked to match with the gaze time. All the gaze points in that time period were used to generate the first feature set. Then, for the second set, the second click to (n+1)-th click were considered to generate the second feature set. This window continuously moved till the last click to generate eye feature sets. This procedure was repeated for eight different window sizes(6-moves, 7-moves, 8-moves, 10-moves, 12-moves, 15-moves, 20-moves, 30-moves) to get multiple overlapping feature sets that capture eye-movement patterns at different temporal resolutions. For generating the eye features, MATLAB 24a was used. Figure 4 shows how the sliding window method was used to generate the features, and Figure 3 shows the flow of how the dataset was prepared and then cleaned for further analysis.

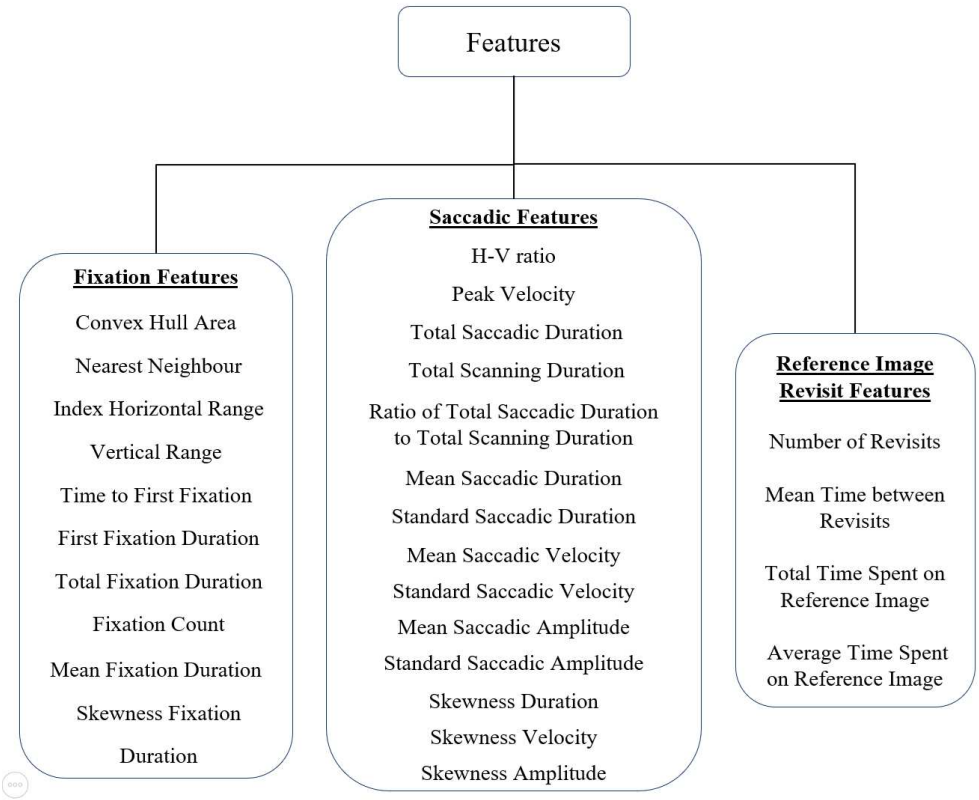


Fig. 2: Gaze Features

3.2.1 Categorization and Split of Participants

This study intended to compare the gameplay behavior of the 'solvers' (who solved the puzzle) and 'Non-solvers' (who could not solve the puzzle). We further analyze the behavioral differences between the subcategories, that is, two types of 'Solvers' – (a) 'Experts' (players who solved the puzzle with a lesser number of moves) and (b) 'Persistent players'(players who solved the puzzle with a greater number of moves), and two types of 'Non-solvers' – (a) 'Triers' (players who could not solve the puzzle but played for long time) and (b) 'Quitters' (players who could not solve the puzzle and left the game early). Hence, the train-test sets were created using those participants' subcategories. Based on their solvability behavior and the duration of gameplay, all 98 participants were divided into 4 sub-categories. The number of moves was used to measure how long participants played the game, and the median of total moves (141 moves) was used to divide participants into groups. This resulted in 24 Experts, 25 Persistent Players, 24 Triers, and 25 Quitters. Train-test participants were selected randomly within their respective sub-categories. As an 80-20 split was used, the training set consisted of 20 participants from each subgroup, while the remaining

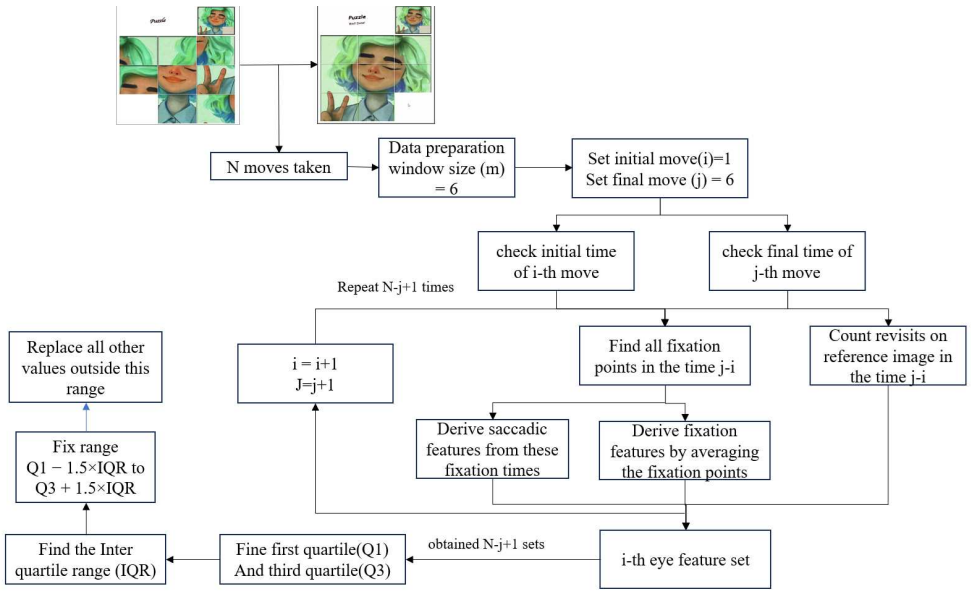


Fig. 3: Data Preparation Process

participants formed the test set. Furthermore, this study validated the result using 10 cross-validation sets.

3.3 Implementation Details

Susac et al. (2019) showed a significant difference between the 'Solvers' and 'Non-solvers', where 'Solvers' have a small number of fixations and a smaller fixation duration in particular. (Krebs et al. 2021) also found significant difference between the Solver's gaze pattern and the Non-solver's gaze pattern. Building on their findings, our current study employed an ensemble-based machine learning classifier (namely, Random-Forest) to automatically differentiate solver and non-solver gaze behaviors. The Random-Forest model was implemented using the Scikit-learn library in Python. This study tuned four Random-Forest hyper-parameters, namely `n_estimators`, `max_depth`, `min_sample_leaf` and `min_sample_split` using the grid-search technique. Random-Forest algorithm has many other hyper-parameters for which we used the default values set by Scikit learn, namely `sqrt` (square-root of all features) for `max_features` and `gini` for criterion. The `sqrt` setting indicates that if the model has m input features, each decision tree is trained using at most $\sqrt{m}$ features at each split. For the cross validation, we created the train-test sets randomly using the participant subcategory. As the gameplay behavior of each participant differs from that of the others, k-fold cross-validation was not used.

A distinct approach was adopted for model evaluation: during training, the entire dataset was used following the conventional procedure, whereas during testing, each participant was evaluated separately. The final model accuracy was then obtained by

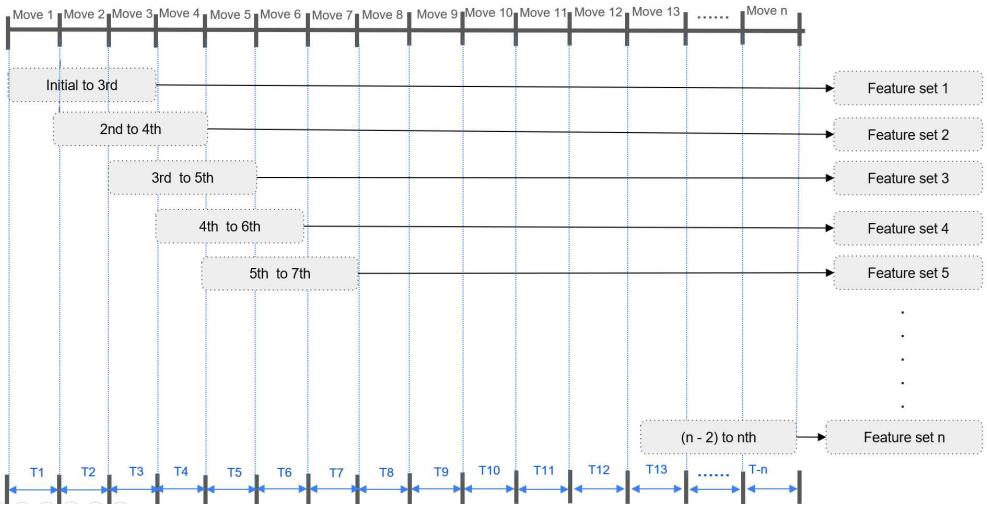


Fig. 4: Representation of an Example Eye-Feature Set Generation using Sliding Window Method (with a Window Size of 3)

averaging the classification accuracies across all participants, providing a more individualistic and strong assessment of model performance. Another evaluation method involved a progressive measure, that is, for every i-th iteration, the model predicted the result for the first i eye-features, and the value of i increased at every iteration.

To predict the remaining gameplay behavior, this study used the moves, as they capture the dynamic progression of problem-solving. To the best of our knowledge, no previous work has employed this approach to model gameplay progression. For this purpose, a Random-Forest Regressor was implemented, and the model was trained and tested using four distinct strategies to explore different behavioral generalisations;

1. **Trained on 'Solvers' + 'Non-solvers' AND Tested on 'Solvers' + 'Non-solvers':** This strategy evaluates the overall generalizability of the model across various participant categories, ensuring both the successful and unsuccessful gameplay patterns are represented.

2.(a) **Trained on 'Solvers' AND Tested on 'Solvers':** This strategy learned the successful gameplay behavior. While testing, the model evaluated in how many steps the 'Solver' would be able to solve.

(b) **Trained on 'Solvers' AND Tested on 'Non-solvers':** In this strategy, while testing, the model evaluated if a 'Non-solver' would solve the puzzle, in how many steps he would be able to solve.

3. **Trained on 'Non-solvers' AND Tested on 'Non-solvers':** This strategy focused on capturing inefficient or unsuccessful problem-solving patterns, assessed how consistently the model could represent struggling participants' gameplay dynamics. This strategy focused on evaluating the time after which someone is going to leave the game. In this strategy, the model was not tested on 'Solvers', as they finished the game and did not have the tendency to quit.

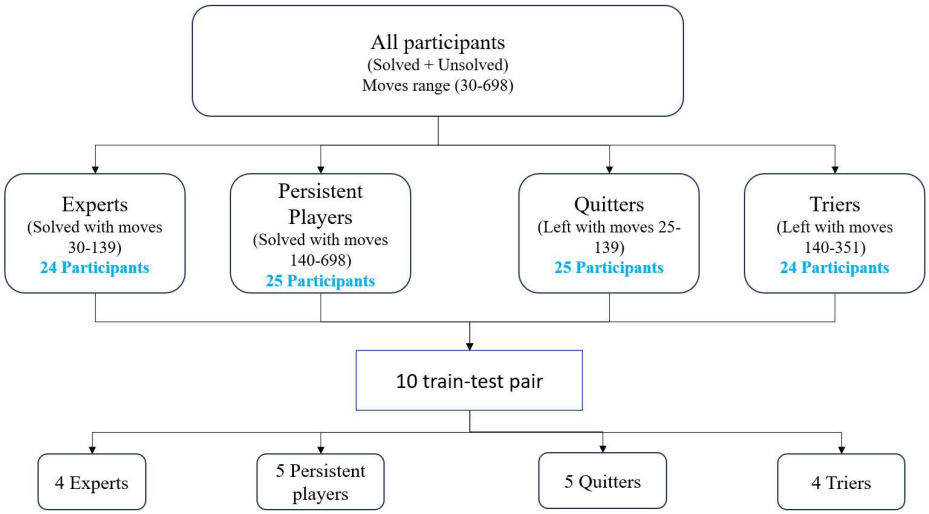


Fig. 5: Demonstration of Participant Categories and Splits based on Solvability Behavior and Required Number of Moves

4. **Trained and Tested within Subcategories:** This strategy aimed to preserve intra-group behavioral characteristics, such as differences between participants who solved or failed with more or fewer moves, allowing the model to capture fine-grained distinctions in problem-solving efficiency.

All the experimental results for the experiments explained in this section, are given in Section 4.

4 Experimental Results and Observations

[7] This section presents the results obtained with Random-Forest classifier and Random-Forest regressor obtained with the gaze behavior.

4.1 Classification of 'Solvers' and 'Non-solvers'

This study implemented Random-Forest classifier using 31 eye-tracking features as input (mentioned in Table A1), and the output represented the participant's solvability behavior. The 'Solver' labels were encoded as 1, and 'Non-solvers' as 0. As mentioned earlier, the classifier was trained on the training set and tested on each participant individually. To obtain an overall performance measure, the classification accuracies from all individual participants were averaged. Figure 6 depicts the overall accuracies across all 10 test sets for each data preparation window. The overall classification

[7] All results obtained with different methodologies are available in the link: https://github.com/Dipshikha-P/All_results

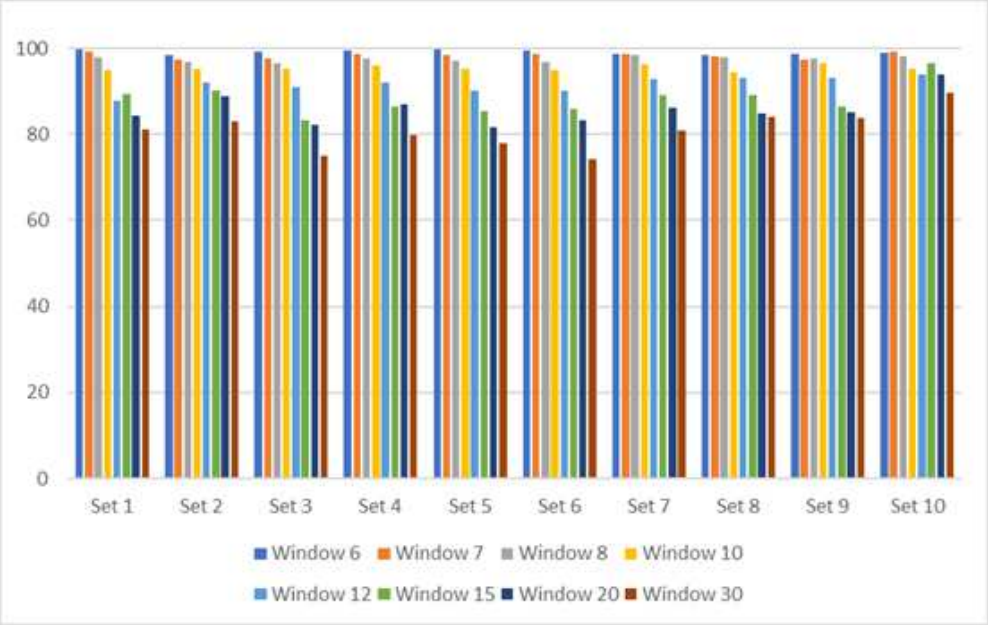


Fig. 6: Prediction Accuracy Obtained by Random-Forest Classifier for all Testsets Generated with Possible Sliding Windows Variations

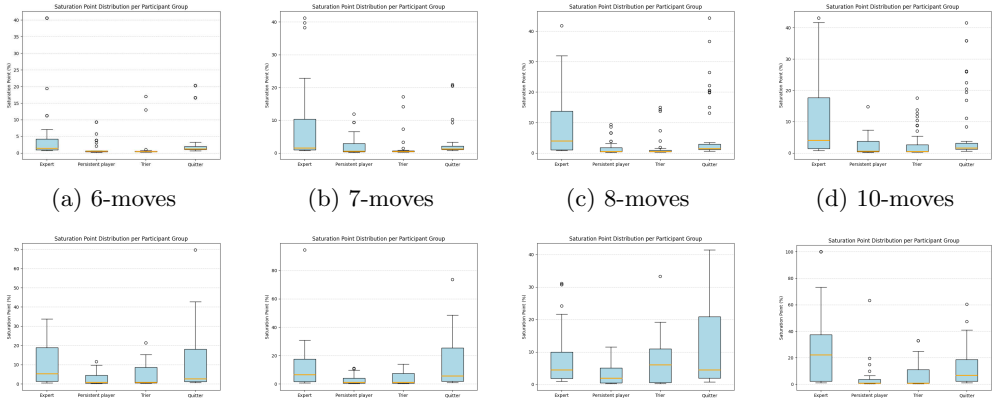
accuracy ranged between 75% and 100%. Notably, for all test sets, the dataset prepared with sliding window sizes between 6 and 10, the average classification accuracy exceeded 90%, demonstrating the model's strong discriminative capability. For window sizes between 12 and 20, the average accuracy remained consistently higher than 80%, indicating that the model could still reliably differentiate 'Solvers' from 'Non-solvers' even with coarser temporal aggregation. In contrast, for some test sets with a sliding window size of 30, the accuracy occasionally dropped below 80%, suggesting that excessively large windows smooth out critical temporal details in the gaze data that are essential for behavioral distinction. This study considered window size 6-30, as a smaller or larger window size would cause data loss. As the optimum number of moves to solve the puzzle was 30, a bigger window size failed to capture the temporal behavior of gameplay. On the other hand, in the case of a smaller window size (i.e., 1-5 moves) it failed to capture some of the gaze features, such as convex-hull-area, and revisit features.

Figure 6 further shows a general decreasing trend in accuracy with increasing window size. However, participant-wise analysis revealed that this decline was not strictly monotonic – for some participants, smaller windows resulted in lower accuracy, whereas larger windows increased it, and vice versa. This variation reflects the diversity in participants' gameplay behavior. Participants who completed the puzzle in fewer moves exhibited rapid strategic shifts; therefore, smaller window sizes captured these local behavioral changes more effectively. Conversely, participants who required more moves demonstrated gradual strategic adaptations, a larger window size averaged out

the noise in their gameplay and captured overarching gaze patterns more efficiently. Overall, the result obtained with Random-Forest classifier demonstrated that gaze behavior alone can predict the problem-solving success, and Random-Forest classifier can effectively learn and generalize these behavioral distinctions.[8]

4.2 Detection of Earliest Saddle-Point

From the result obtained with Random-Forest classifier, it was found that the classifier can distinguish 'Solvers' and 'Non-solvers' with very high accuracy – for some participants, it reaches 100% classification accuracy. This suggests that the model is capable of identifying the behavioral differences between the two groups from a very early stage of the game. To investigate how early this differentiation occurs, the classifier was evaluated on the test data progressively across the gameplay. The move window after which the classifier's accuracy did not decrease suddenly or only increased was defined as the *saddle point*. In other words, this point represented the stage of gameplay after which the model's predictions stabilize, indicating that the player's behavioral pattern has become distinct and consistent. For quantitative analysis, the saddle point was expressed as a percentage of total moves, i.e., the move number at which the models' accuracy stabilized, divided by the participant's total moves. This fractional representation enables comparison across participants with different total move lengths.



(a) 6-moves (b) 7-moves (c) 8-moves (d) 10-moves

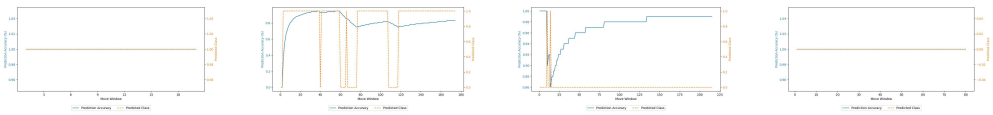
(e) 12-moves (f) 15-moves (g) 20-moves (h) 30-moves

Fig. 7: Box Plots Obtained with Saddle Points for all Test Participants Derived with Various Sliding Windows

A global analysis of these saddle points provides insights into how quickly each subcategory of participants (Experts, Persistent Players, Triers, and Quitters) settled

[8]Individual classification accuracy for all test sets obtained with the dataset prepared with an 8-move window is given in the appendix B.1, and the results obtained with datasets prepared with other data preparation windows are attached in the following link: https://github.com/Dipshikha-P/All_results/tree/main/Indivisual_classification_graphs

into their characteristic strategies. An individual-level analysis offered a deeper understanding by revealing how early each of the models' predictions stabilized for each participant within their respective group. Figure 7 shows the box plots of the percentage values of saddle points obtained for all test participants across different sliding window sizes[9]. It was observed that for datasets prepared using window sizes between 6 and 12, the classifier reached stability within the first 5% of total moves for most participant groups — including Persistent Players, Triers, and Quitters. For Experts, this percentage appeared slightly higher (around 20%). It happened due to the lower number of total moves; the individual plots show that the model's prediction stabilized at an early stage for these participants, just like others. But after normalizing, the percentage value increases relative to others. For larger windows (15 to 30), the saddle point appeared between 5–40% of total moves for most groups except Persistent Players. The apparently higher percentage saddle point in larger window datasets occurs due to the reduced number of data samples caused by window averaging — even if the model stabilizes within 20–30 moves, the normalized percentage value appears to be larger. The finding of this subsection indicates that the solver and non-solver behavioral differences emerge at a very early stage of gameplay, often within the first few moves. Figure 8 shows the individual classification plot of one participant from each subcategory.



(a) Expert (b) Persistent (c) Trier (d) Quitter

Fig. 8: Individual Classification Plot based on Sub-Categories

4.3 Prediction of Remaining Game-Play Behavior

As the classifier successfully differentiated 'Solvers' and 'Non-solvers' at an early stage, this study further checked whether it is possible to predict the remaining gameplay for which we predicted remaining moves based on the gaze data and the number of moves already taken. The Random-Forest Regressor was used for this prediction with four different strategies, namely – (a) Train on all participants, where the model learned the general behavior; (b) Train on only solved participants, where the model learned efficient goal-directed behavior; (c) Train on only unsolved participants, where the model learns error-prone inefficient behavior; and (d)Train on participant subcategories, where the model was trained of group specific behavior. To determine which strategy best predicted the remaining gameplay, the Mean Absolute Error (MAE) metric was used for evaluation. The predicted remaining moves were then plotted against the actual number of moves taken.

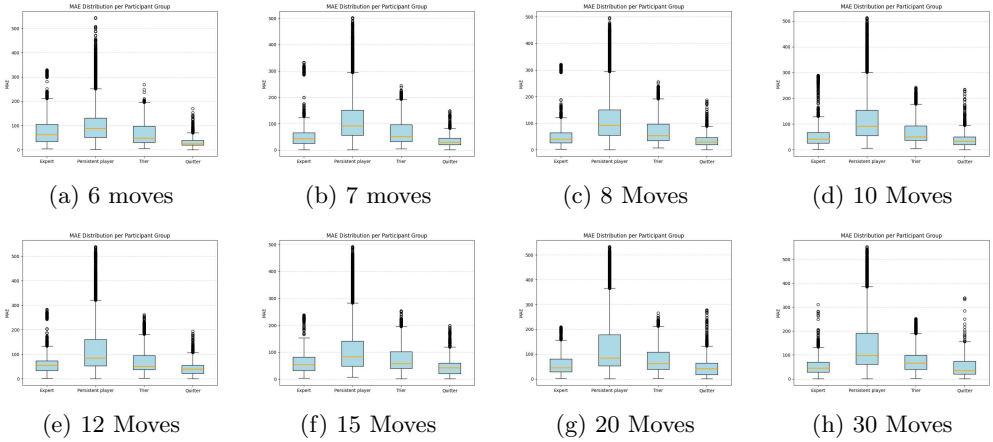
[9]The scattered plots obtained with the individual saddle point values are given in the link: https://github.com/Dipshikha-P/All_results/tree/main/saddle%20point%20detection

After that, the predicted remaining moves were plotted against the actual moves taken. In the ideal scenario, the predicted remaining moves should decrease continuously, giving us a diagonal line. But human gameplay is noisy (Graham et al. 2023), they fall into loops, they take wrong moves, then by backtracking they undo the wrong moves or change strategy based on the trap, hence we obtained graphs with an erratic pattern. Sometimes the moves increased or decreased continuously with small noise (Dörner and Güss 2022). A global analysis of these saddle points provides insights into how quickly each subcategory of participants (Experts, Persistent Players, Triers, and Quitters) settled into their characteristic strategies. An individual-level analysis offered a deeper understanding by revealing how early each of the models' predictions stabilized for each participant within their respective group. Figure 7 shows the box plots of the percentage values of saddle points obtained for all test participants across different sliding window sizes[10]. It was observed that for datasets prepared using window sizes between 6 and 12, the classifier reached stability within the first 5% of total moves for most participant groups — including Persistent Players, Triers, and Quitters. For Experts, this percentage appeared slightly higher (around 20%). It happened due to the lower number of total moves; the individual plots show that the model's prediction stabilized at an early stage for these participants, just like others. But after normalizing, the percentage value increases relative to others. For larger windows (15 to 30), the saddle point appeared between 5–40% of total moves for most groups except Persistent Players. The apparently higher percentage saddle point in larger window datasets occurs due to the reduced number of data samples caused by window averaging — even if the model stabilizes within 20–30 moves, the normalized percentage value appears to be larger. The finding of this subsection indicates that the solver and non-solver behavioral differences emerge at a very early stage of gameplay, often within the first few moves. Figure 8 shows the individual classification plot of one participant from each subcategory.

4.3.1 Optimal Model Training among Four Strategies

As mentioned before, Random-Forest regressor was trained and tested with different groups of people following different strategies. To identify which strategy predicts the gameplay behavior better, the mean absolute error(MAE) was calculated for all 18 participants in the test set, prepared with different data preparation windows. Figures 9, 10, 11, 12 show the summary of MAE values obtained for each participant with different sliding window sizes used in data preparation. Based on the MAE summaries, Random-Forest regressor trained on subcategories predicted the remaining gameplay better than all other Random-Forest regressors, as the MAE values for all subgroups are lower than the MAE values obtained with other strategies. The logic behind this result is that, in this strategy, the model learns specific group behavior, hence in predicting the remaining moves, the difference between the actual move and the predicted move is not huge. As shown in Figure 13 and 14 an 'Expert' participant took 60 moves, which the model trained on the subgroup predicts to be 57, whereas

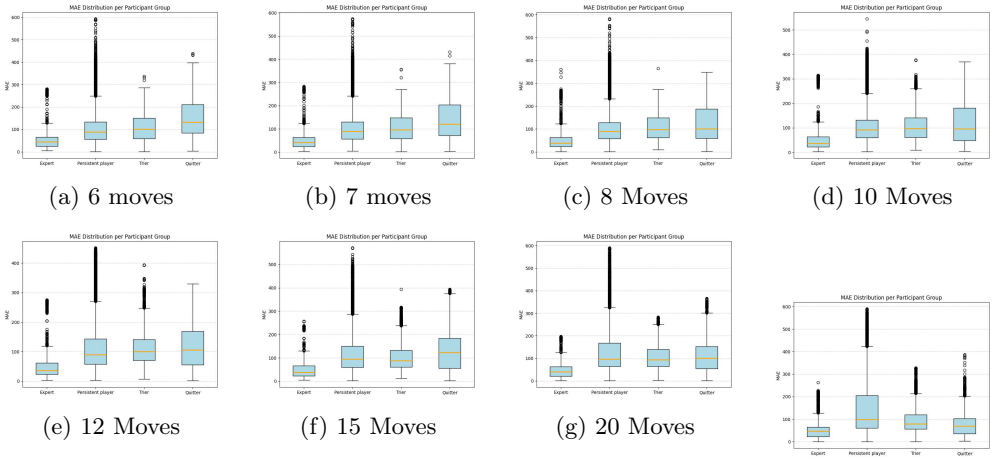
[10]The scattered plots obtained with the individual saddle point values are given in the link: https://github.com/Dipshikha-P/All_results/tree/main/saddle%20point%20detection



(a) 6 moves (b) 7 moves (c) 8 Moves (d) 10 Moves

(e) 12 Moves (f) 15 Moves (g) 20 Moves (h) 30 Moves

Fig. 9: Representation of Mean Absolute Error Obtained for Each Participant when Model was Trained on all Participants

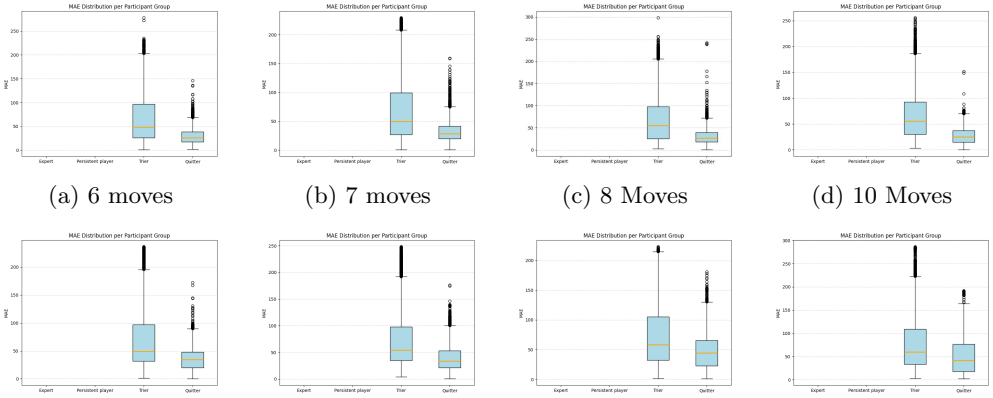


(a) 6 moves (b) 7 moves (c) 8 Moves (d) 10 Moves

(e) 12 Moves (f) 15 Moves (g) 20 Moves

Fig. 10: Representation of Mean Absolute Error Obtained for Each Participant when Model was Trained on 'Solvers'

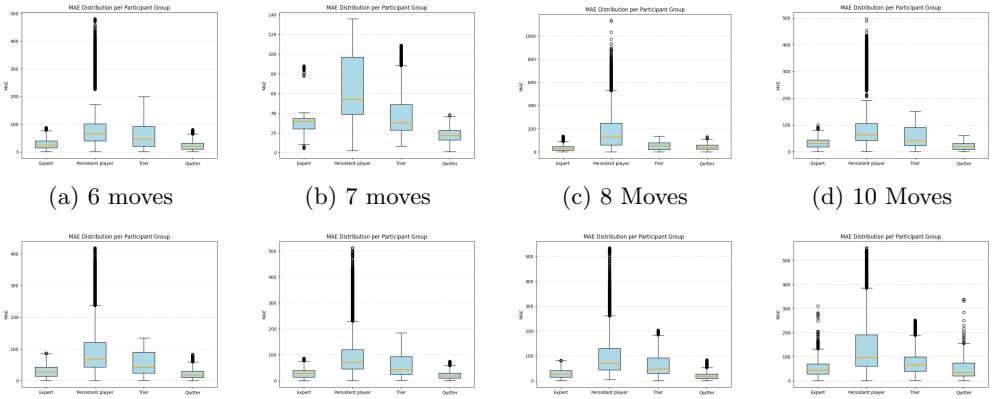
the model trained on all participants predicted it to be 140. This behavior of the models was also found in the other participants. Figures 9 and 10 indicate that when the model was trained on all participants, it performed relatively well because the model learned the generalized behavioral pattern encompassing both 'Solvers' and 'Non-solvers', which helped it generalize better to unseen data. However, when the model was trained only on 'Solvers' or only on 'Non-solvers', there was a higher chance of overfitting, as the model learned only the solver or the non-solver behavior. Figure 10 shows that when the model was trained on 'Solvers' and tested on 'Non-solvers', its



(a) 6 moves (b) 7 moves (c) 8 Moves (d) 10 Moves

(e) 12 Moves (f) 15 Moves (g) 20 Moves (h) 30 Moves

Fig. 11: Representation of Mean Absolute Error Obtained for Each Participant when Model was Trained on 'Non-solvers'



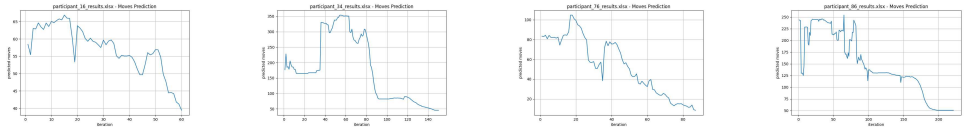
(a) 6 moves (b) 7 moves (c) 8 Moves (d) 10 Moves

(e) 12 Moves (f) 15 Moves (g) 20 Moves (h) 30 Moves

Fig. 12: Representation of Mean Absolute Error Obtained for Each Participant when Model was Trained on Subcategories, i.e., 'Experts', 'Persistent players', 'Triers' and 'Quitters' Separately

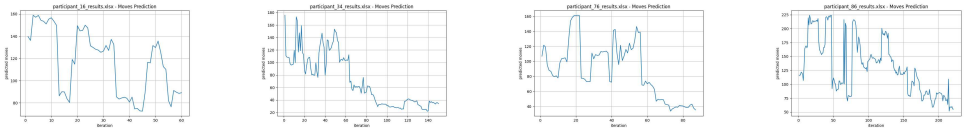
performance dropped significantly, as expected, since the learned and observed behaviors were opposite in nature. Interestingly, the model predicted the behavior of 'Triers' better than that of 'Quitters'. 'Triers' attempted to solve the puzzle and exhibited solver-like behaviors during certain stages, resulting in lower MAE values compared to 'Quitters'. From these results, it can be concluded that the most effective strategy for predicting remaining gameplay behavior is to train the model on subcategories, followed by training on all participants as the second-best approach. However, when the predicted remaining moves were analyzed individually, it was observed that the

overall graphical patterns for different training strategies were mostly similar. The participant-level results are discussed in 4.3.2.



(a) Expert (b) Persistent (c) Quitter (d) Trier

Fig. 13: Individual Regression Graphs of Four Participants from each Subcategory Obtained with Random-Forest Regressor Trained on Subcategories



(a) Expert (b) Persistent (c) Quitter (d) Trier

Fig. 14: Individual Regression Graphs of Four Participants from each Subcategory Obtained with Random-Forest Regressor Trained on all Participants

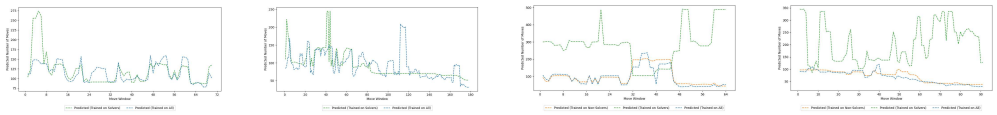
Although the model trained on individual subcategories produced the lowest prediction error, we used the results obtained from Random-Forest regressor trained on all participants for further analysis. Figure 13 and 14 indicate that, though the model trained on subcategories captured the intrinsic behavioral pattern of a specific group, it failed to reflect the behavioral fluctuation in the gameplay. On the contrary, the model trained on all participants produced a noisier trajectory, indicating it successfully captured the variability in the dataset. The model trained on subcategories learned the stable, group-level characteristics due to the homogeneity of its training data. Whereas the model trained on all participants was exposed to a diverse behavioral pattern. Hence, it was more sensitive to the behavioral variability of the dataset.

4.3.2 Predicting Individual Behavior

To gain a deeper understanding of how Random-Forest Regressor captured the progression of gameplay, the predicted remaining moves were plotted (on the y-axis) against the moves already taken (on the x-axis). Ideally, the predicted remaining moves should decrease steadily as the gameplay advances. However, the plots revealed a continuous fluctuating pattern. Figure 15 presents four graphs corresponding to each participant category.

For the 'Expert' and 'Persistent players', two models were compared in the graph—one trained on all participants and another trained only on 'Solvers'. The

'Trier' and 'Quitter' graphs include an additional model trained on 'Non-solvers'. It can be observed that, during the initial phase of gameplay, both strategies for 'Solvers' exhibit an erratic pattern. However, towards the end, the model trained only on 'Solvers' produces a smoother prediction curve. This occurs because near to the completion of the puzzle, participants' search behavior becomes more structured and stable, leading to more consistent gaze patterns. Consequently, the solver-only model learns less behavioral variability and produces smoother predictions. On the other hand, the model trained on both 'Solvers' and 'Non-solvers' was exposed to a wider range of behavioral variations, enabling it to analyze gaze transitions more thoroughly and resulting in more fluctuations in its predictions.



(a) Expert (b) Persistent player (c) Quitter (d) Trier

Fig. 15: Individual Graphs for 'Experts', 'Persistent Players', 'Quitters' and 'Triers' to Visualize Predicted gameplay Behavior

For 'Non-solvers', both trier and quitter graphs show that when the model was trained on all participants or only on 'Non-solvers', the predicted remaining moves decreased over time. In contrast, when the model was trained on 'Solvers', the predicted remaining moves tend to increase. This indicates that the model trained on 'Solvers' learned efficient strategic behavior – the opposite of what is typically observed in 'Non-solvers' – resulting in increasing predicted moves for inefficient gameplay. Moreover, in the Quitter's graph, this model presents an upward trend at the end, suggesting that the participant was moving away from the solution. In contrast, the 'Trier' graph shows a decreasing trend towards the end, indicating that some 'Triers' were actually approaching the solution but discontinued. Individual analyzes also revealed that for participants who solved the puzzle in fewer than 40 moves, Random-Forest Regressor was unable to accurately predict the remaining moves. This suggests that the model requires a certain number of gameplay observations to understand the participant's gaze behavior and stabilize its predictions.

4.4 Characterization of Group behavior

The individual outputs from Random-Forest classifier and regressor were examined to gain a better understanding of how different gameplay strategies and problem-solving behaviors emerged among participants. This analysis established the presence of different behavioral patterns in the groups found by both the classifier and regressor. The behavioral categorizations obtained from the classification and regression analyzes are presented in the following subsections, emphasizing how each model revealed distinct facets of 'Solver' and 'Non-solver' performance dynamics.

4.4.1 categorization based on Results Obtained with Random Forest Classifier

After analyzing the results of Random-Forest classifier with different test sets, we found three distinct result patterns. Figure 8 has four individual plots obtained after progressively testing Random-Forest classifier. To illustrate the results, the individual classification plots obtained with all testsets prepared with data preparation window 8 are given in appendix B.1. The three types of patterns found with Random-Forest classifier are discussed below, and the interpretation is discussed in Section 4.5.

(i) The Random-Forest classifier stabilized at the early stage of gameplay, which was captured through the saddle point, and the overall classification accuracy was high, i.e., 98-100%. This pattern was mostly found in people who took fewer moves, whether to solve or quit.

(ii) The Random-Forest classifier stabilized at an early stage but as the classifier provided wrong predictions of class, the classification accuracy was lower than 98%. This behavior was found in the participants who took a larger number of moves to solve or quit.

(iii) The Random-Forest classifier required some time to stabilize, and then the classification accuracy continued to rise. This behavior was found in the players who took a larger number of moves to solve the puzzle. For a few participants, the classifier demonstrated a small decrease in accuracy in between, giving an overall classification accuracy of around 75-93%.

4.4.2 Categorization based on Results Obtained with Random Forest Regressor

After analyzing the individual predicted gameplay behavior from the remaining moves graphs, it was found that the behavioral difference between 'Solvers' and 'Non-solvers' was best captured by Random-Forest Regressor model trained only on 'Solvers'. The training strategy that used both 'Solvers' and 'Non-solvers' learned a mixture of efficient and inefficient behaviors; as a result, when tested on both types of participants, it was unable to capture any clear behavioral distinction. However, when the model was trained only on 'Solvers', it primarily learned efficient strategic gameplay and, as expected, could accurately predict when a test participant who solved the puzzle was approaching completion. When the same model was tested on 'Non-solver' participants, it struggled to predict the remaining gameplay accurately. This is because 'Non-solvers' demonstrated different behavioral patterns. For some of them, as they approached the end of their gameplay, the difference between the actual remaining moves and the predicted remaining moves increased, indicating a behavioral mismatch between the model's learned strategy and the participants' real actions.

The individual graphs also revealed differences within each group. Figure 15 shows that the predicted gameplay behavior of 'Persistent players' was erratic, compared to that of 'Experts'. Similarly, among 'Non-solvers', 'Triers' showed noisier gameplay patterns than 'Quitters'. 'Experts', specifically those who completed the puzzle within or near the optimal number of moves – tend to follow a strategic and stable gameplay pattern from the beginning, resulting in smoother and less noisy predictions. Conversely,

Table 1: Mean Standard Deviation, Mean Coefficient Variation and Mean Signal-to-Noise Ratio Obtained with Window Size 8 for all 'Solvers' in Various Testsets

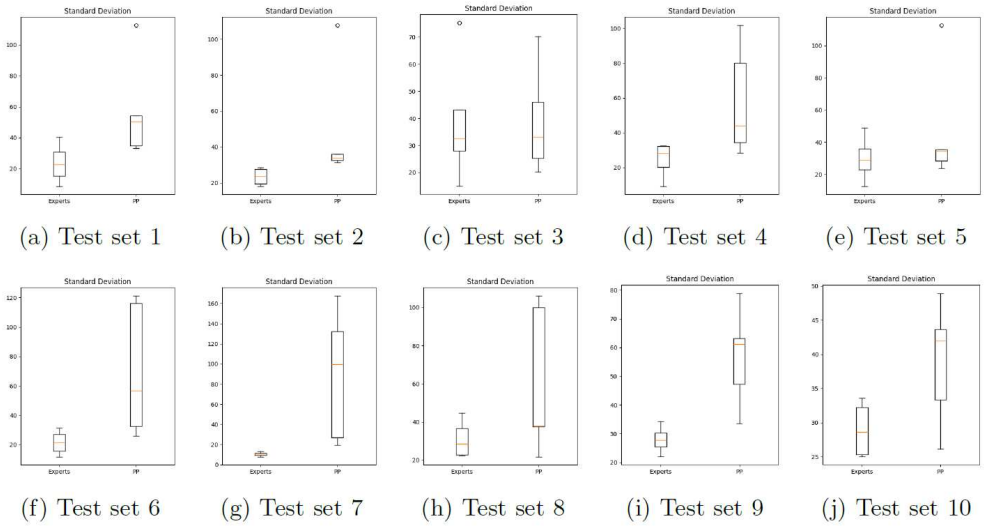
Set	Expert Std	PP Std	Expert CV	PP CV	Expert SNR	PP SNR
Set 1	23.484	56.972	0.233	0.531	5.660	2.014
Set 2	23.437	48.280	0.272	0.434	3.797	2.326
Set 3	25.582	51.708	0.259	0.540	4.253	1.959
Set 4	24.402	57.753	0.274	0.561	6.451	2.012
Set 5	29.790	46.853	0.321	0.473	3.521	2.144
Set 6	21.361	70.522	0.212	0.442	5.017	2.405
Set 7	10.414	89.169	0.108	0.519	9.688	2.412
Set 8	31.013	60.696	0.250	0.555	4.461	2.348
Set 9	27.938	56.816	0.296	0.537	3.879	1.959
Set 10	28.946	38.811	0.181	0.480	6.408	2.168

Table 2: Mean Standard Deviation, Mean Coefficient Variation and Mean Signal-to-Noise Ratio Obtained with Window Size 8 for all 'Non-solvers' in Various Testsets

Set	Trier Std	Quitter Std	Trier CV	Quitter CV	Trier SNR	Quitter SNR
Set 1	50.343	37.848	0.375	0.471	2.836	2.244
Set 2	54.274	34.390	0.564	0.433	1.841	3.088
Set 3	48.386	35.401	0.573	0.450	1.814	2.311
Set 4	57.908	32.228	0.448	0.400	2.251	2.561
Set 5	47.446	39.581	0.415	0.446	2.442	2.494
Set 6	52.330	34.395	0.612	0.494	1.705	2.361
Set 7	42.00	30.604	0.392	0.351	3.093	2.956
Set 8	66.755	33.575	0.488	0.379	2.167	2.975
Set 9	61.985	36.687	0.663	0.416	1.620	2.974
Set 10	59.084	25.681	0.610	0.337	1.686	3.597

'Solvers' who took more moves often experienced false starts, repeated loops, or frequent strategy shifts, which introduced higher variability in their gameplay – a pattern that was reflected as increased noise in their predicted remaining moves. To further analyze these behavioral variations and compare gameplay behavior between experts and persistent players, the standard deviation of predicted moves was computed and represented using box plots for better visualization (Figure 16). The figure shows eight box plots depicting the standard deviation values for 'Experts' and 'Persistent players' in Test Set 1 across all data-preparation windows. Each box plot represents the deviation in predicted moves for individual participants within the group. It was observed that the standard deviation values for 'Experts' were consistently lower than those of 'Persistent players', indicating more stable and consistent behavior.

To examine this observation better, two additional metrics – the Coefficient of Variation (CV) and the Signal-to-Noise Ratio (SNR) – were also calculated. The SNR is an effective indicator of signal quality: a higher SNR value denotes a cleaner, less



(a) Test set 1 (b) Test set 2 (c) Test set 3 (d) Test set 4 (e) Test set 5

(f) Test set 6 (g) Test set 7 (h) Test set 8 (i) Test set 9 (j) Test set 10

Fig. 16: Standard Deviation Value of Predicted Remaining Moves for 'Solvers' in all Testsets Prepared with Data Preparation Window 8

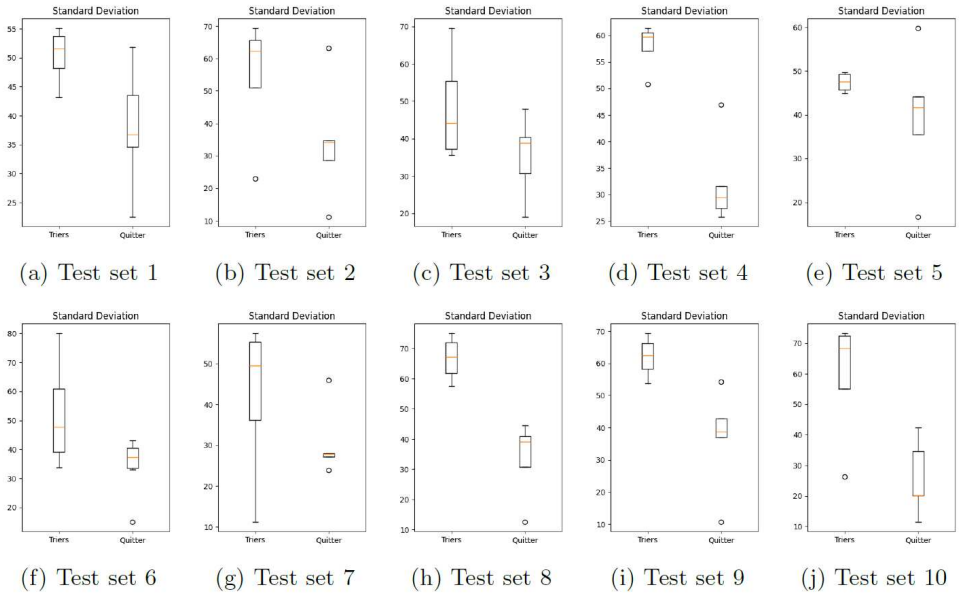
noisy signal, while a lower value indicates higher noise[11] (Wang et al. 2013). The average standard deviation, CV, and SNR values across all test sets obtained with a data-preparation window size 8 moves are summarized in Table 1 for the 'Solvers'. The table shows that the standard deviation value is consistently lower for 'Experts' but higher for 'Persistent players', and the SNR value is consistently higher for 'Experts', but it is lower for 'Persistent players'.

A similar analysis was performed for the non-solver group. It was found that 'Triers' had higher standard deviation values compared to 'Quitters', implying greater noise in their gameplay patterns (Wang et al. 2013). 'Triers' appeared to make genuine attempts to solve the puzzle – occasionally making correct moves and momentarily aligning with solver-like gaze behaviors. This intermittent overlap between efficient and inefficient strategies likely contributed to the higher noise observed in their gameplay. Figure 17 presents the comparative results for the 'Non-solver' group. The SNR value of the 'Non-solvers' indicates a similar result, as the SNR value for the 'Quitters' is higher than that of the 'Triers'.

4.5 Gameplay Behavior and Human Cognition

Existing works show that eye movement affects how we perceive a problem (Turano and Heidenreich 1999; Spivey and Dale 2011) and it also affects the way we solve any problem. Alejandro Lleras mentioned in his article that the 'Non-solvers' fail to solve problems because of how they see the problem (McGaughey 2007). Literature suggests, through gaze behavior, it is possible to assess whether someone takes an efficient strategy to solve a problem or takes an inefficient strategy (Van Gog et al. 2005).

[11] https://en.wikipedia.org/wiki/Standard_deviation



(a) Test set 1 (b) Test set 2 (c) Test set 3 (d) Test set 4 (e) Test set 5

(f) Test set 6 (g) Test set 7 (h) Test set 8 (i) Test set 9 (j) Test set 10

Fig. 17: Standard Deviation Value of Predicted Remaining Moves for 'Non-Solvers' in all Testsets with Data Preparation Window 8

Hence, this study analyzed the gaze features mentioned in Table A1 to understand the problem-solving behavior. To analyze the gameplay behavior of participants, we have used the following methods:

1. Individual analysis across gameplay
   (a) Plotting of an individual's accuracy over gameplay. We conducted this for all 18 participants in the test set.
   (b) Plotting individual saddle points as a percentage of total moves.
   (c) Plotting of an individual's predicted remaining moves over gameplay, which generated 18 plots for each participant in each test set.[12]
2. Group analysis of gameplay behavior where the participants were divided into 'Solvers' and 'Non-solvers'.

The individual graphs obtained with the classification results showed that for some of the 'Non-solvers', the classification accuracy was throughout 100% whereas the classification accuracy for most of the 'Solvers' was less than 100% and had a dynamic behavior. According to literature, the change of behavior of the machine learning models occurs due to the change in behavioral parameters used to train the models (in this study, it is the gaze behavior) (Yarkoni and Westfall 2017; Kahou et al. 2013). Hence, based on the results from Random-Forest classifier, we can conclude

[12] All individual plots obtained with both Random-Forest Classifier and Random-Forest Regressor are given in appendix B.1 and B.2 for one data preparation window (8 moves window).The plots obtained with other data preparation windows are available in the followinglink: https://github.com/Dipshikha-P/All_results

that, the non-solvers were following a different path than the optimal path (To elaborate this finding the individual classification graphs of the 'Non-solvers' are given in appendix B.1 which shows that most of the 'Non-solvers' had a throughout 100% classification accuracy) and the 'Solvers' were continuously changing their strategy to reach the goal (To elaborate this finding, the individual classification graph of the 'Solvers' are shown in appendix B.1 which depicts that the classification accuracy of these participants were changing throughout the gameplay.)

The classification summary of the average group behavior shows that the classification accuracy is more than 90% (Figure 6), and as the result of Random-Forest classifier was a reflection of the gaze behavior (Yarkoni and Westfall 2017; Kahou et al. 2013), indicating that the 'Solvers' and 'Non-solvers' had significant differences in gaze behavior. The gaze behavior continuously changes in the exploratory phase, and as a player finds out the efficient path to solve a problem, it stabilizes (Knoblich et al. 2001; Tsai et al. 2012). Hence, to identify when the solvers settled down in their solving strategy, we used Random-Forest classifier progressively while testing the model. The point where the model prediction settled down is defined as the saddle point in the Section 4.2, and it also indicates that the participant's gaze behavior also settled down at that stage, which indicates where the participant is going to click on (Land et al. 1999; Land and Hayhoe 2001). Unique behavior patterns of an individual cluster of participants were observed from the saddle points analyzes and the individual classification accuracy. We have attempted to label these behaviors separately for convenience.

- Participants who solved the game and selected the optimal pathway to the solution at the very beginning: The Random-Forest classifier gave high accuracy around 98-100%, and the model prediction settled very early. These participants could be identified as the 'Experts' (Figure 8a).
- Participants who solved the problem with a delayed time and settled down with the optimal pathway a little late (i.e., false start): The Random-Forest classifier gave a moderate accuracy around 75-93% and the model prediction settled after few moves. These participants could be identified as 'Persistent players' (Figure 8b).
- Participants who approached the problem through a longer path but later optimized their behavior, yet did not complete the solution: The Random-Forest classifier gave an accuracy around 80-90% but the model prediction settled at an early stage. These participants could be identified as 'Triers' (Figure 8c)
- Participants who did not solve the problem: The Random-Forest classifier gave high accuracy around 98-100%, and the model prediction settled very early. These participants could be identified as the 'Quitters' (Figure 8d).

Findings from the analysis of the number of moves remaining to complete or quit the task also found two distinct behaviors in the 'Solvers'. One had a smoother graph for the remaining number of moves, and the other group had a graph that was continuously fluctuating. This behavior was also supported by standard deviation, CV value and SNR value[13]. As already mentioned, the change in the machine learning model's

[13]The standard deviation value, CV value and SNR value obtained for the dataset prepared with a window size 8-moves are given in Table 1 and the box plots obtained with the standard deviation value for the same dataset are given in Figure 16. These values for the other datasets are given in the appendix C,

result occurs due to the change in the behavior parameter (Yarkoni and Westfall 2017; Kahou et al. 2013), from the results of the remaining gameplay prediction, it can be said that the second group had a noisier gaze behavior compared to the first group. Literature shows that when someone inferred the solution path or knows a solution to a problem, their gaze behavior changes systematically land2006eye, based on this theory, the first group can be labelled as the 'Experts'. It is also found that when a person searches for the solution, he shows a noisier and reactive gaze behavior (Sailer et al. 2005), based on this theory the second group can be labelled as the 'Persistent players'. Both the groups that solved the puzzle displayed a rapid reduction of predicted moves to solve the problem after a period of game play. This indicates that they have developed some 'insight' into solving the problem at hand (Knoblich et al. 2001).

Findings from the analysis of the number of moves remaining to quit the task reflect that the 'Non-solvers' can also be divided into two groups, one had a smoother graph, and another had an unstable or noisier graph. This behavior was also supported by standard deviation, CV value and SNR value (Table 2). Previous studies suggest that if a person has no clear idea about the solution path of a problem but intends to solve it, he shows continuous variation in his behavior (Graf et al. 2023), on the contrary, individuals with no intention to solve a problem do not shift their approach after getting stuck (Wilson et al. 2014). Based on these theories, the first group can be labelled as the 'Triers' and the second group can be labelled as the 'Quitters'.

Random-Forest regressor indicates that several participants in the 'Triers' group had reached the optimal pathway to solve the problem near the end of their gameplay behavior, as was reflected by a rapid reduction of the number of moves predicted to complete the task. However, these participants quit before completion. This implies that they did not realize that they were on the right path. This gives a hint of the absence of insight in the gameplay behavior. Other participants in the 'Trier' group showed a rise in the number of moves predicted to complete the task, indicating that they were on an inefficient path when they quit the game. A similar behavior was observed for the quitters.

5 Conclusion

The aim of this study was to investigate the role of gaze behavior in understanding the problem-solving behavior for which the ensemble-based machine learning model, Random Forest model, was trained on the eye tracker dataset. The results demonstrate that the gaze dynamics can effectively capture problem-solving and gameplay behavior of an individual. The Random Forest classifier and regressor results were combined to reveal unique participant behavioral patterns. The classifier results showed that although 'Non-solvers' displayed more stable patterns in their gameplay behavior, 'Solvers' displayed constant oscillations. This implies that while 'Non-solvers' exhibited monotonous and unchanging behavior, 'Solvers' constantly modified their approaches and adjusted to new solutions. The study also identified transitions from exploratory

and the boxplots for other datasets are available in the following link: https://github.com/Dipshikha-P/All_results/tree/main/Standard%20deviation%20of%20remaining%20moves

to stable behavior using the saddle point concept, which denotes the point at which participants found the best course of action.

This study further investigated the behaviors of the 'Solvers' and 'Non-solvers' groups and found two more distinctions between both of the groups, i.e., 'Experts', 'Persistent players' in the 'Solver' group and, 'Triers' and 'Quitters' in the 'Non-solver' group. The behavioral distinction between the groups could be captured by both the classification model and the regression model. This study also successfully detected the moment when the solvers found an insight into the solution. Interestingly, some 'Triers' also demonstrated signs of insight into the solution, which is an important indicator for educators because any type of hint at this point may help the 'Trier' to solve the problem. On the contrary, the 'Quitters' did not find any insight into the gameplay.

Though this study successfully captured the behavioral distinctions between the four subcategories, it was found that few participants in one subcategory reflected a behavior pattern like the other subcategory. For example, very few 'Quitters' developed some insight to solve the problem towards the end of the game, as depicted by the individual regression plots of the Quitters, which were mainly found in 'Triers'. Also, a few of the 'Experts' reflected a fluctuation in their gaze behavior as captured by the individual classification graphs of the 'Experts'. Moreover, a few Persistent players showed consistency in their solvency behavior. It suggests that the subgroups obtained with the results and the subgroups we created through game solvability behavior and total number of moves were not the same, indicating the number of moves alone is not a reliable marker for the categorization. For instance, one person may purposefully choose a longer route but make effective movements at every stage, while another may begin with an ineffective move but swiftly adjust their approach to get a comparable or even shorter solution path Gray and Fu (2004).

The findings of this study confirm the notion that gaze behavior is a sensitive indicator of cognitive strategy and problem-solving purpose. The used paradigm, which combines machine learning and eye tracking, provides a fresh method for objectively differentiating between exploratory, adaptive, and disengaged behaviors in challenging problem-solving situations. This approach can be used to investigate the relationship between visual attention patterns and the effectiveness of thinking, learning, and decision-making in other cognitive activities.

Declaration

- **Funding:** The authors thank the Department of Science and Technology, New Delhi, India for funding this study under the Cognitive Science Research Initiative Scheme.
- **Authors' contribution:** Conceptualization: AH, RG, PPC; Methodology: DP, AH, RG; Software: DP; Validation: RG, AH; Investigation: DP, SA; Writing-Original Draft: DP; Writing- Review and Editing: DP, AH, RG, PPC; Visualization: DP; Supervision: AH, RG, PPC; Project administration: AH, RG, PPC; Funding acquisition: RG

Appendix A Feature Extraction Equations

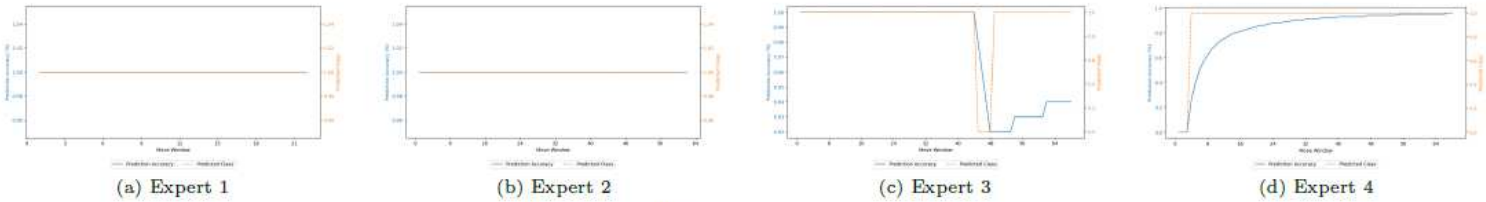
The equations used in this study to extract different gaze features are given in Table A1

Table A1: Feature extraction equations

Feature	Extraction Formula
Convex Hull Area	$A = \frac{1}{2}\sum_{i=1}^{n}(x_i y_{i+1} - y_i x_{i+1})$, Here $(x_{n+1}, y_{n+1}) = (x_1, y_1)$
Nearest Neighbour Index	$NNI = \frac{\frac{1}{n}\sum_{i=1}^{n} d_i}{\frac{2}{n}\sqrt{\frac{A}{n}}}$, d_i=distance to the nearest other point in fixation i
Horizontal Range	$HR = x_{max} - x_{min}$
Vertical Range	$VR = y_{max} - y_{min}$
First Fixation Duration	$D_{fix_1} = final_{fix_1} - init_{fix_1}$
Total Fixation Duration	$\sum_{i=1}^{n} D_i$, n = Number of Fixations
Mean Fixation Duration	$x' = \frac{1}{n}\sum_{i=1}^{n} D_i$
Skewness Fixation Duration	$\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\frac{(x_i - x')^3}{S}$, S = Standard Deviation
H-V ratio	$\frac{StimulusHeight}{StimulusWidth} \cdot \frac{\sum_{i=1}^{n} x_{i+1} - xi}{\sum_{i=1}^{n} y_{i+1} - y_i}$
Peak Velocity	$V_i^{\circ}/sec = 2.tan^{-1}(\frac{Amp_i^{pix}}{T_i/3600})$, $Amp_i^{pix} = \sqrt{(x_i - y_{i+1})^2 + y_i - x_{i+1})^2}$ 1° of visual angle = 3600 pixels V_j is the j^{th} Saccadic Velocity, the paper considered 3 largest saccadic velocities
Total Saccadic Duration	$\sum_{i=1}^{n} D_{Sacc_i} = \sum_{i=1}^{n} init_{fix_{i+1}} - (init_{fix_i} + D_{fix_i})$
Ratio to Total Saccadic Duration to Total Scanning Duration	$\frac{\sum_{i=1}^{n} D_{Sacc_i}}{TotalScanningDuration}$
Mean Saccadic Duration	$D_{Sacc_{Mean}} = \frac{\sum_{i=1}^{n} D_{Sacc_i}}{n}$, Mean Saccadic Velocity and Mean Saccadic Amplitude were calculated using a similar equation
Standard Saccadic Duration	$D_{Sacc_{Std}} = \sqrt{\frac{\sum_{i=1}^{n}(D_{Sacc_i} - D_{Sacc_{mean}})^2}{n-1}}$, Standard Saccadic Velocity and Standard Saccadic Amplitude were calculated using a similar equation

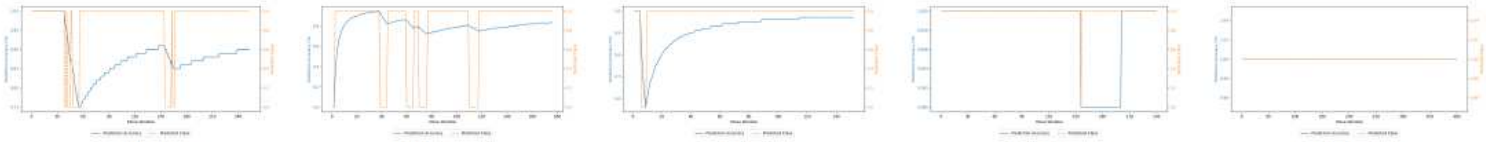
Appendix B Additional figures

B.1 Individual Classification Accuracy



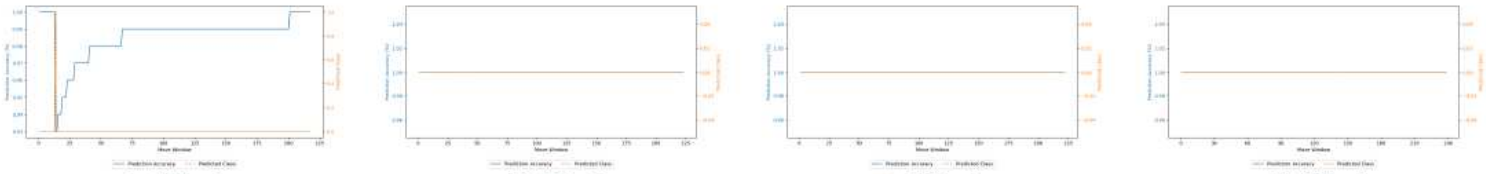
(a) Expert 1 (b) Expert 2 (c) Expert 3 (d) Expert 4

(a) Individual classification accuracy plot of all experts



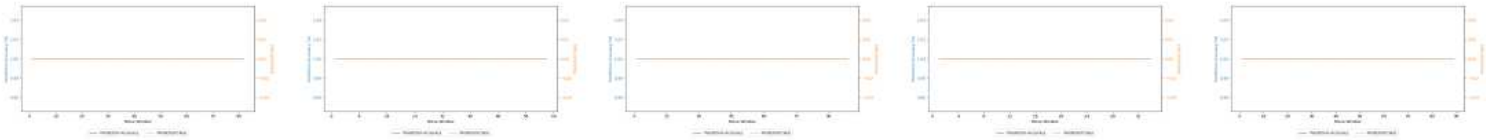
(a) Player 1 (b) Player 2 (c) Player 3 (d) Player 4 (e) Player 5

(b) Individual classification accuracy of all Persistent players



(a) Trier 1 (b) Trier 2 (c) Trier 3 (d) Trier 4

(c) Individual classification accuracy of all Triers



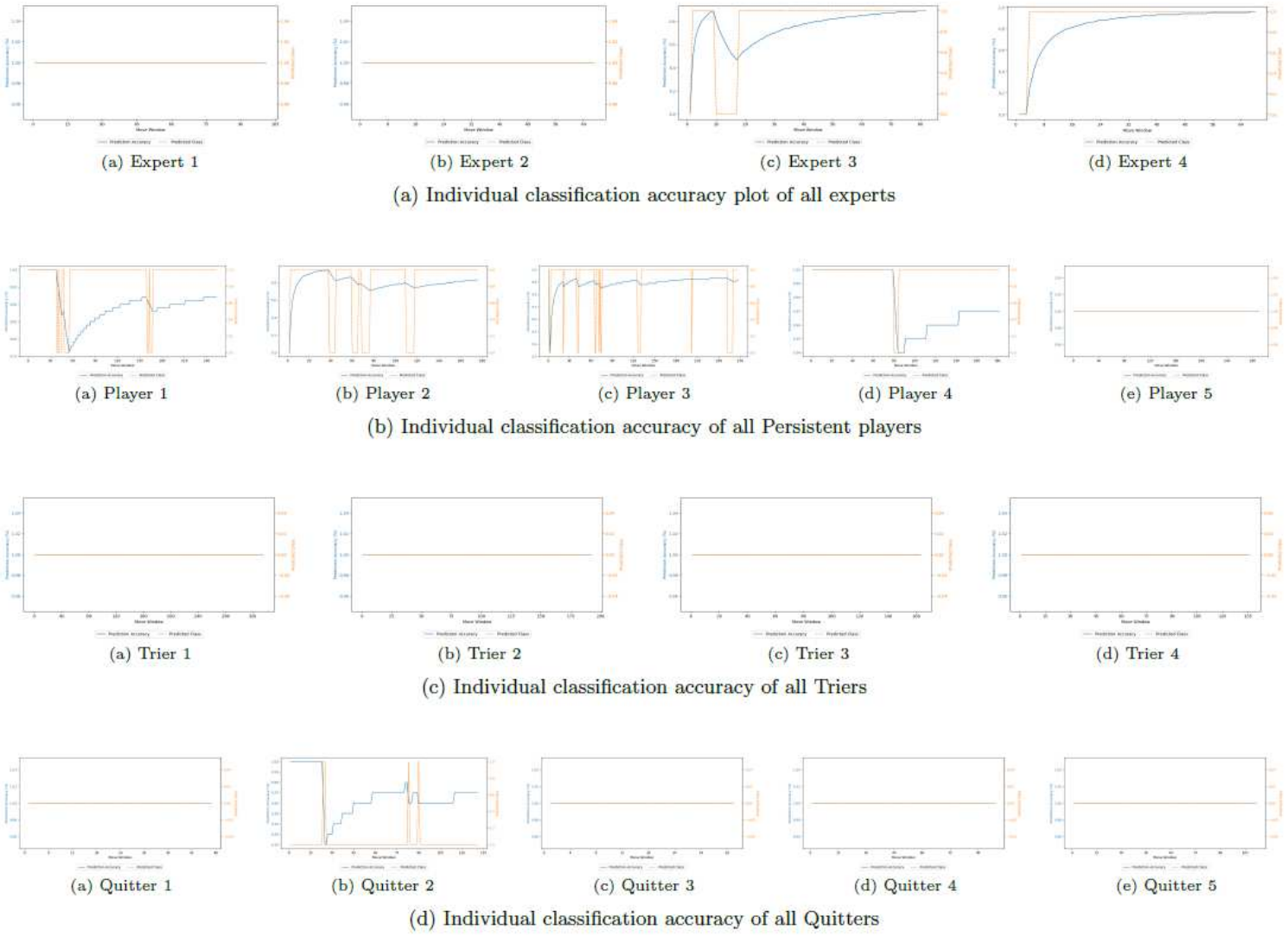
(a) Quitter 1 (b) Quitter 2 (c) Quitter 3 (d) Quitter 4 (e) Quitter 5

(d) Individual classification accuracy of all Quitters

Fig. B1: Individual classification accuracy of all participants in Testset 1 where the orange line indicates the predicted class and the blue line indicates the classification accuracy

Figures B1 to B10 represent the classification graphs obtained with different test sets. Figure B11 depicts that two 'Experts' in Testset 1 obtained a 100% classification accuracy throughout, which indicates that Random-Forest classifier could successfully identify the player's solvability behavior throughout the gameplay. In the case of the third 'Expert', the classifier gave wrong predictions for some times, but the prediction settled at an early stage, which is shown with the orange line. In the case of the fourth 'Expert', the saddle point was obtained after a few moves in the gameplay. In the case of two 'Persistent' players, the prediction

accuracy was 100% and the classifier's prediction settled at an early stage. The saddle point was obtained after some moves for two other 'Persistent players,' and the classifier gave a wrong prediction in between the gameplay for one of those participants, which indicates that the participants had different gaze behavior than the optimum behavior at these stages. In case of player 1 in the 'Persistent' group, the saddle point was found early, but the classifier gave a wrong prediction in between the gameplay, which caused the accuracy drop. This indicates that the participant took correct moves at the initial stage, but after some time, he took some wrong moves, which decreased Random-Forest model's classification accuracy. Figure B1 shows that, except for one player, the classifier identified the solvability behavior of all the 'Non-solvers'.



(a) Expert 1 (b) Expert 2 (c) Expert 3 (d) Expert 4

(a) Individual classification accuracy plot of all experts

(a) Player 1 (b) Player 2 (c) Player 3 (d) Player 4 (e) Player 5

(b) Individual classification accuracy of all Persistent players

(a) Trier 1 (b) Trier 2 (c) Trier 3 (d) Trier 4

(c) Individual classification accuracy of all Triers

(a) Quitter 1 (b) Quitter 2 (c) Quitter 3 (d) Quitter 4 (e) Quitter 5

(d) Individual classification accuracy of all Quitters

Fig. B2: Individual classification accuracy of all participants in Testset 2 where the orange line indicates the predicted class and the blue line indicates the classification accuracy

The Random-Forest classifier gave a similar result as Testset 1 for Testset 2, which is shown in Figure B2. For two 'Experts', it produced 100% classification accuracy. In the case of two 'Persistent' players, the saddle point was found at an early stage, but the

accuracy dropped after some time as the players had different gaze behavior than that of a 'Solver'. The figure shows that, similar to Testset 1, the classifier had 100% classification accuracy for all 'Non-solvers' except one 'Quitter'. This finding indicates that the gaze behavior of the 'Non-solvers' in this set was similar to the expected 'Non-solver' behavior. A 'Quitter' in this group showed some fluctuation in the prediction, which indicates that, sometimes, the 'Quitters' may have gaze behavior similar to the 'Solvers'.

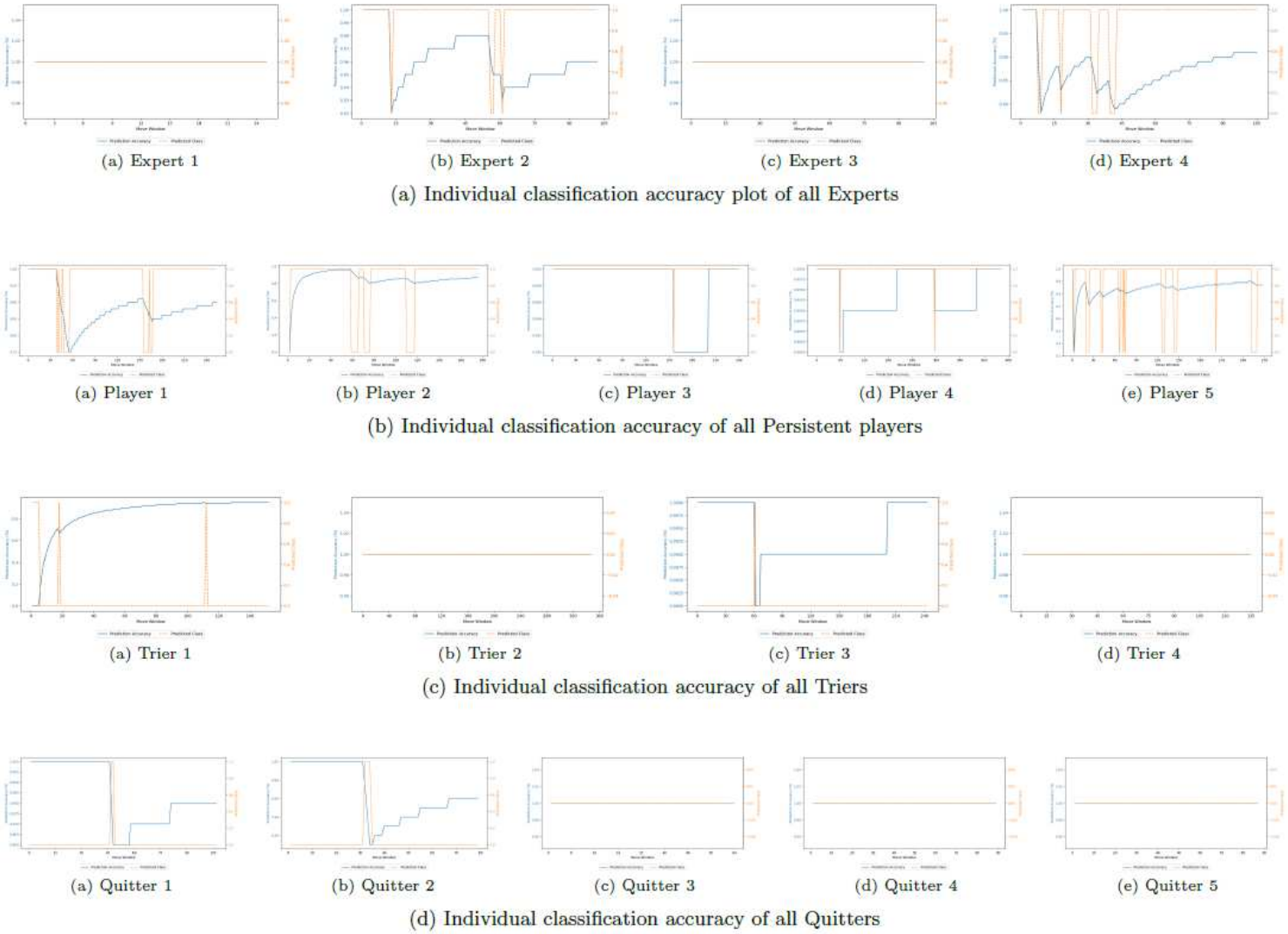


Fig. B3: Individual classification accuracy of all participants in Testset 3 where the orange line indicates the predicted class and the blue line indicates the classification accuracy

The results obtained with Testset 3 are shown in Figure B3, which depicts a similar behavior in 'Solvers' like Testset 1 and Testset 2. However, in this testset the 'Non-solvers' showed a different accuracy graph than the other two testsets. One 'Trier' in this group had classification accuracy around 80%, and the other three had classification accuracy around 100%. In the 'Quitter' group, two participants had an accuracy less than 100% because, though the saddle point was found in an initial stage, the classifier predicted the class incorrectly for some time in between the gameplay.

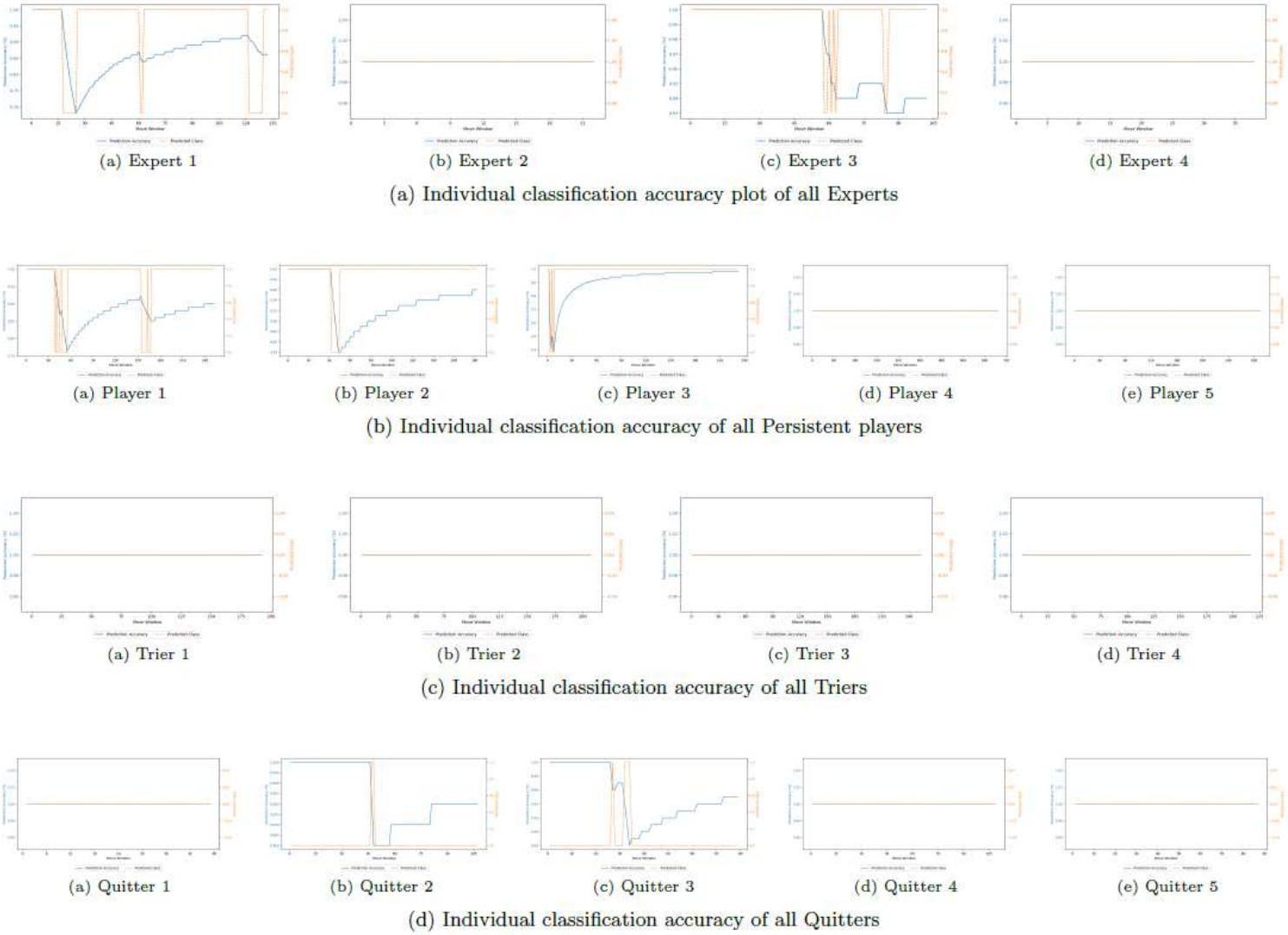
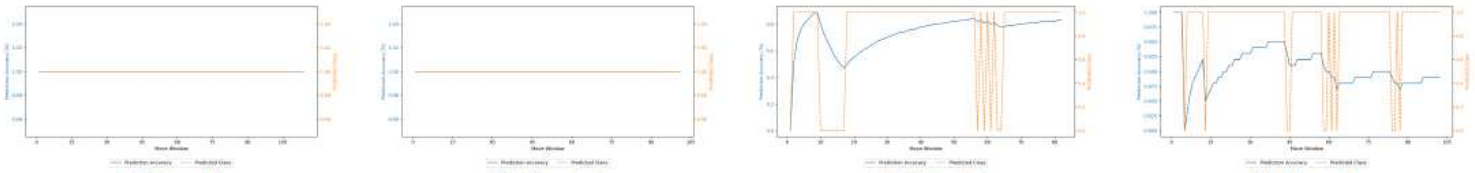


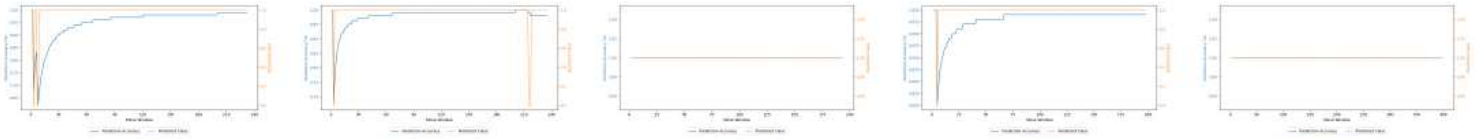
Fig. B4: Individual classification accuracy of all participants in Testset 4 where the orange line indicates the predicted class and the blue line indicates the classification accuracy

Figure B4 shows the changes in classification accuracy throughout the game, obtained with Testset 4. For three 'Persistent' players and one 'Expert' in this group, the saddle point was found after a few moves in the gameplay; for the rest of the 'Solvers', they settled in their gameplay behavior at a very early stage, as the saddle point was found at early stage. Except for two 'Quitters' in this group, for all 'Non-solvers' the classification accuracy was 100%, which indicates that these Non-solvers' gaze behavior was steady throughout and similar to the expected 'Non-solver' behavior.



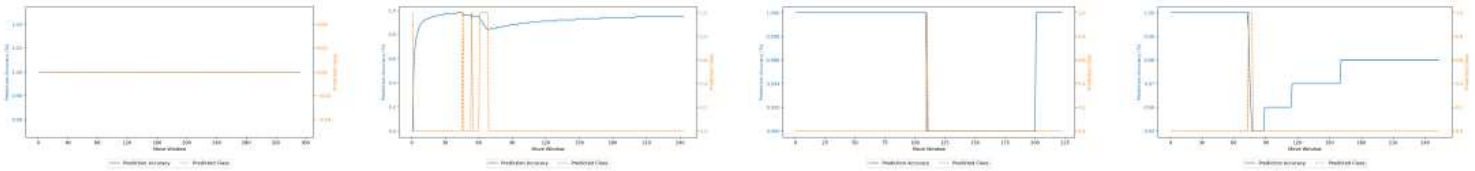
(a) Expert 1 (b) Expert 2 (c) Expert 3 (d) Expert 4

(a) Individual classification accuracy plot of all Experts



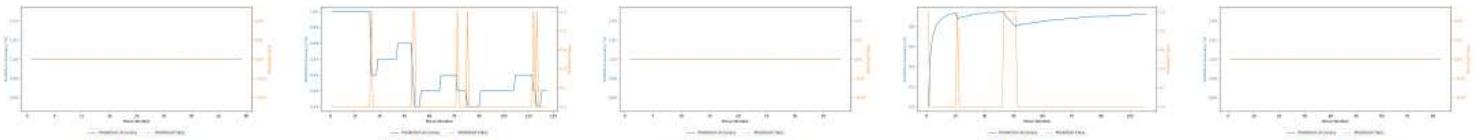
(a) Player 1 (b) Player 2 (c) Player 3 (d) Player 4 (e) Player 5

(b) Individual classification accuracy of all Persistent players



(a) Trier 1 (b) Trier 2 (c) Trier 3 (d) Trier 4

(c) Individual classification accuracy of all Triers



(a) Quitter 1 (b) Quitter 2 (c) Quitter 3 (d) Quitter 4 (e) Quitter 5

(d) Individual classification accuracy of all Quitters

Fig. B5: Individual classification accuracy of all participants in Testset 5 where the orange line indicates the predicted class and the blue line indicates the classification accuracy

In the case of Testset 5 (Figure B5), all the 'Solvers' except two had higher classification accuracy, more than 95%, two 'Persistent' players and two 'Experts' found the saddle point after a few steps, which indicates that the gaze behavior of these participants settled after playing the puzzle for some times. Among these four participants, the classification accuracy of two 'Experts' fluctuated for a few moments in between the game, which indicated that at those moments, the gaze behavior also changed, meaning the gameplay behavior also changed. Among the 'Non-solvers', the classification accuracy of three 'Quitters' and one 'Trier' was less than 100%. For all of these participants, the saddle point was found early, but Random-Forest model predicted the class incorrectly between the game play for some times.

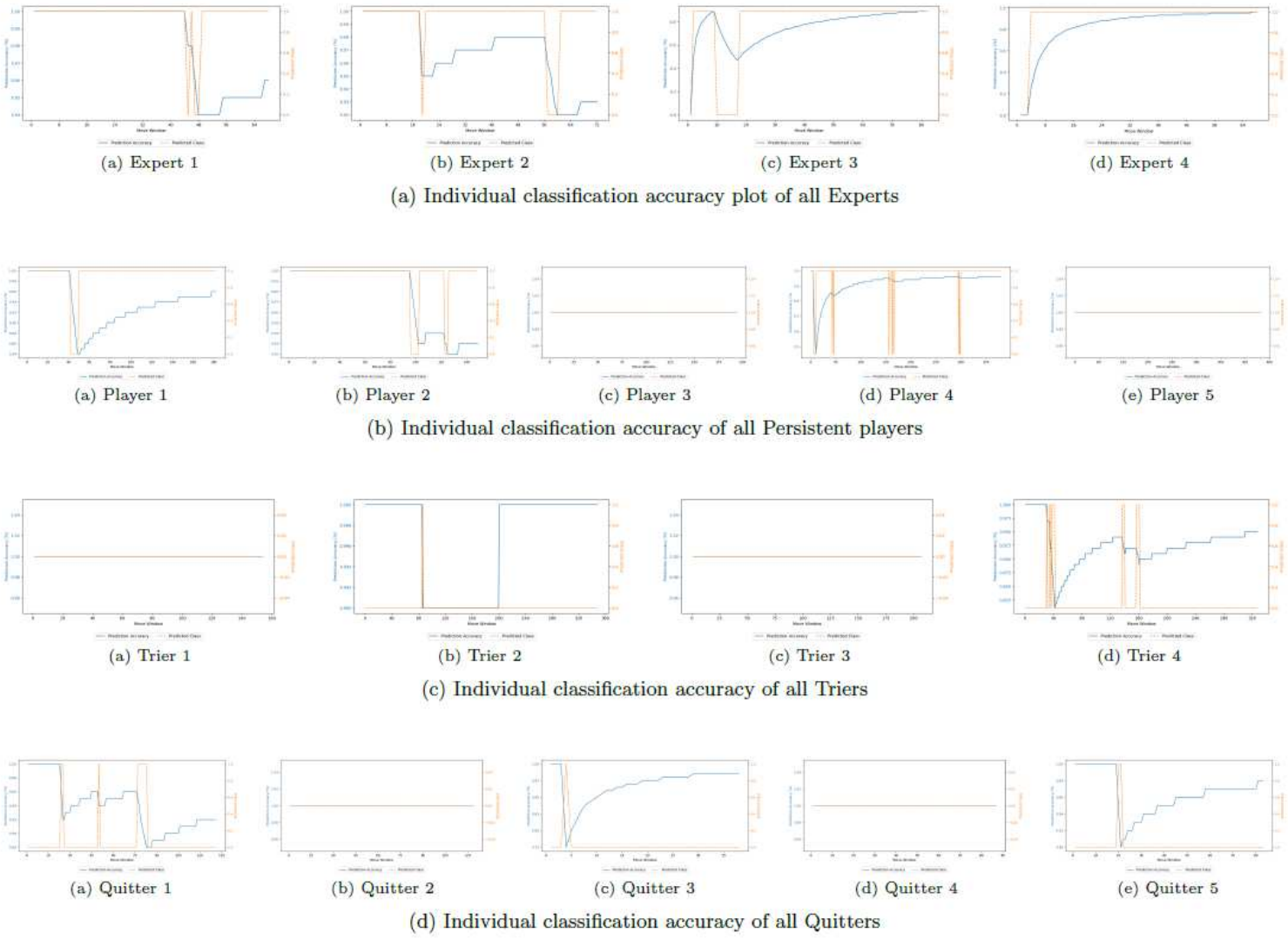


Fig. B6: Individual classification accuracy of all participants in Testset 6 where the orange line indicates the predicted class and the blue line indicates the classification accuracy

Figure B6 shows that in the case of Testset 6, all 'Experts' had a classification accuracy lower than 100%, for two of them the saddle point was obtained early, and for the other two of them the saddle point was obtained after a few moves. This figure shows that three of the 'Persistent' group had lower classification accuracy, and their saddle point was found after a few moves in gameplay. For three 'Triers' and two 'Quitters', Random-Forest classifier obtained 100% classification accuracy throughout gameplay. Except for two 'Non-solvers', all the 'Non-solvers' settled down in their gameplay behavior at an early stage of gameplay.

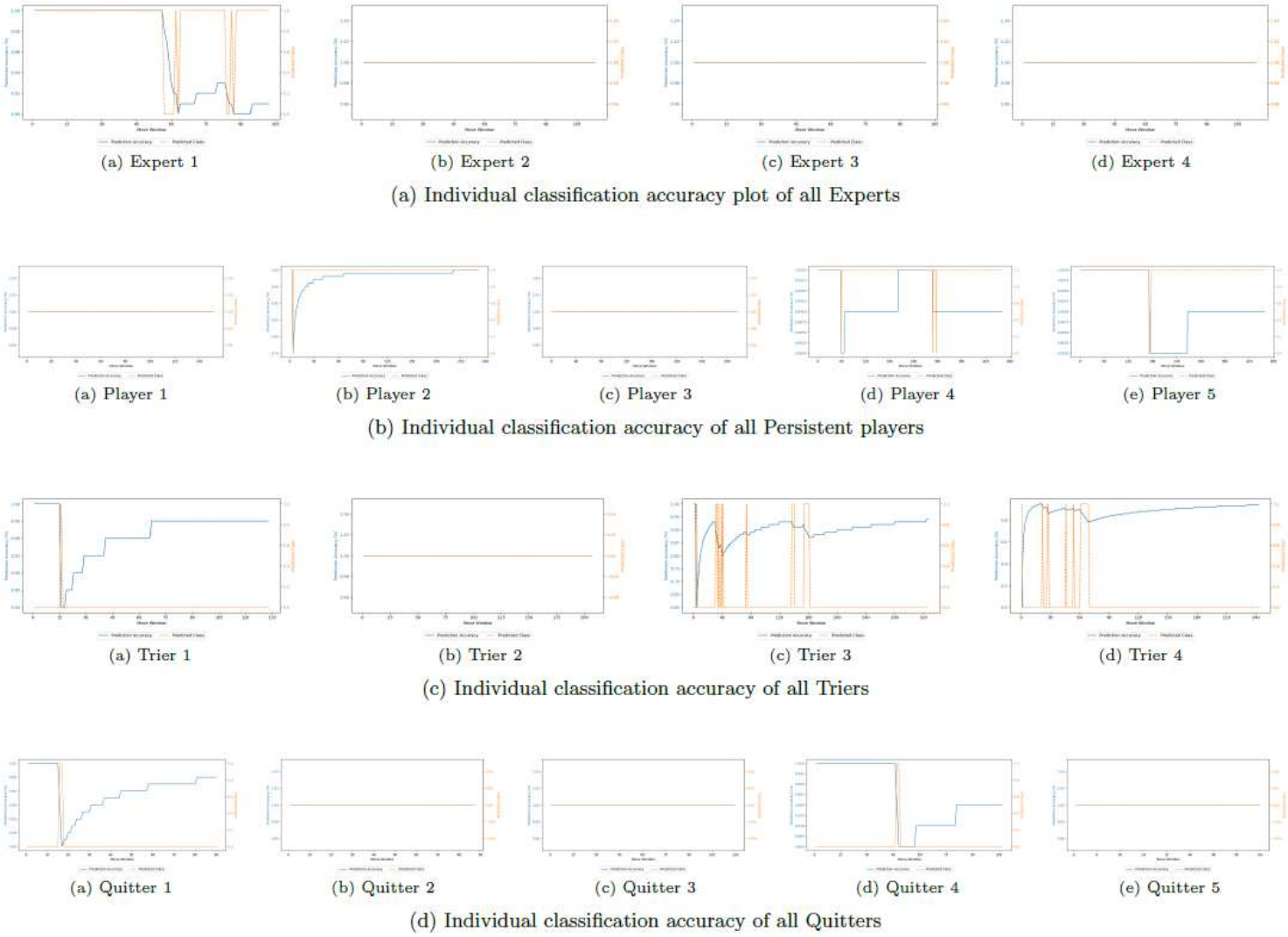


Fig. B7: Individual classification accuracy of all participants in Testset 7 where the orange line indicates the predicted class and the blue line indicates the classification accuracy

Figure B7 has the accuracy graph obtained with Testset 7, which was prepared with data preparation window 8. It depicts that all 'Solvers' in this set had classification accuracy more than 90% and all these participants settled in their gameplay at an early stage. For one 'Expert' in this group, the classifier gave wrong predictions for some time, which indicates that at that stage, the 'Expert' had gaze behavior different from the other 'Experts' in the training group. Among the 'Non-solvers' in this group, for one 'Trier' and three 'Quitters', Random-Forest model obtained a 100% accuracy; other 'Quitters' and one 'Trier' had classification accuracy between 95% and 100%. On the other hand, in the case of the other 'Triers', the classifier gave wrong predictions in between, as a result the classification accuracy was lower than 95%.

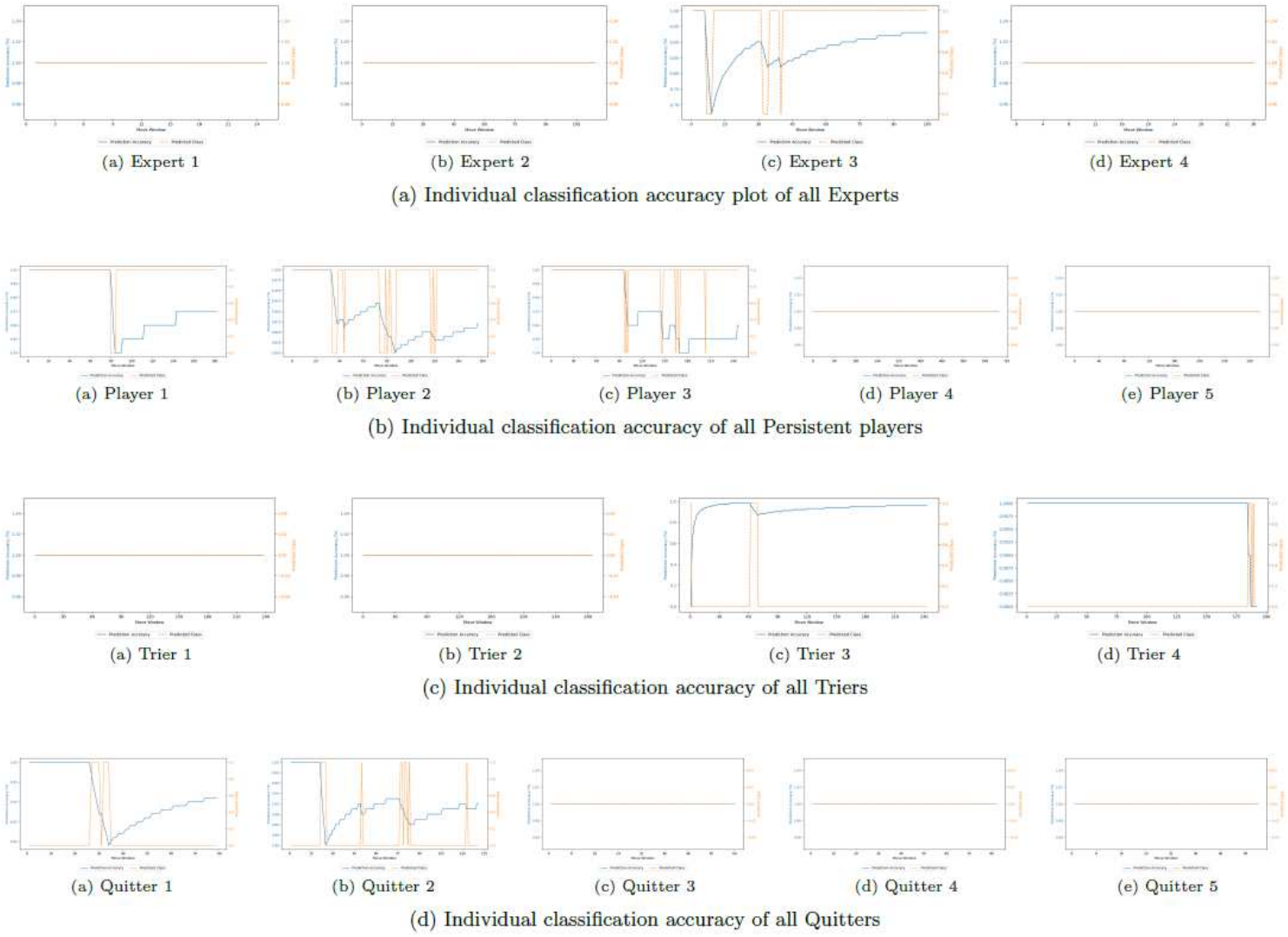


Fig. B8: Individual classification accuracy of all participants in Testset 8 where the orange line indicates the predicted class and the blue line indicates the classification accuracy

Figure B8 shows the classification accuracy of different participants in Testset 8. For three of the 'Experts' and two of the 'Persistent' players, the obtained accuracy was 100% classification accuracy. The other 'Experts' settled in his gameplay after spending a few time in the game which is indicated by a late saddle point. In the case of the other three 'Persistent' players, the players settled down in their gameplay at an early stage, but the classifier prediction fluctuated in between, which indicates that during the gameplay their gaze behavior fluctuated. Among the 'Non-solvers', six players had a steady gaze behavior throughout the gameplay, which was indicated by their 100% classification accuracy.

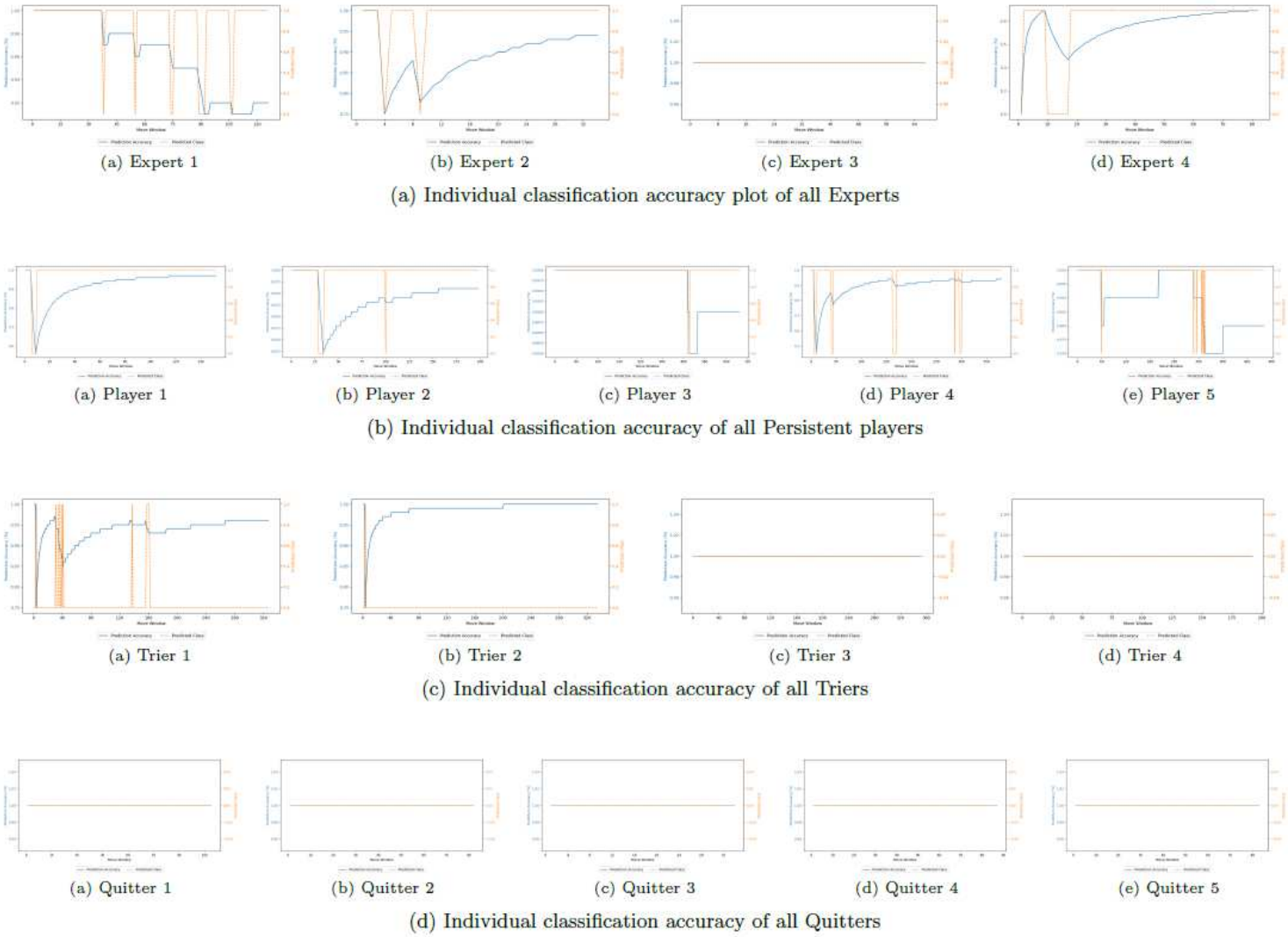
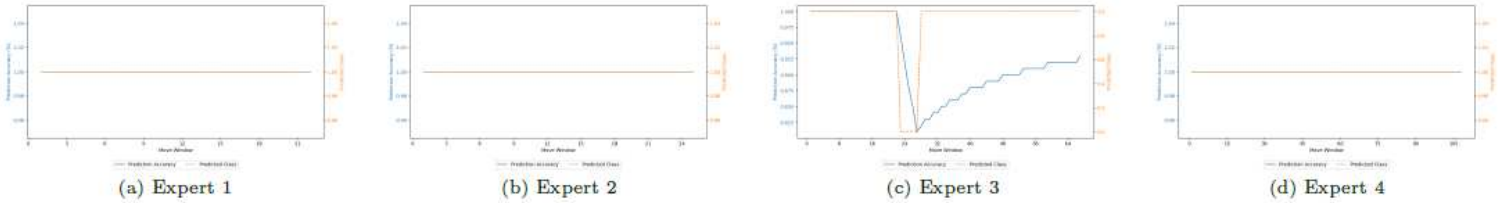


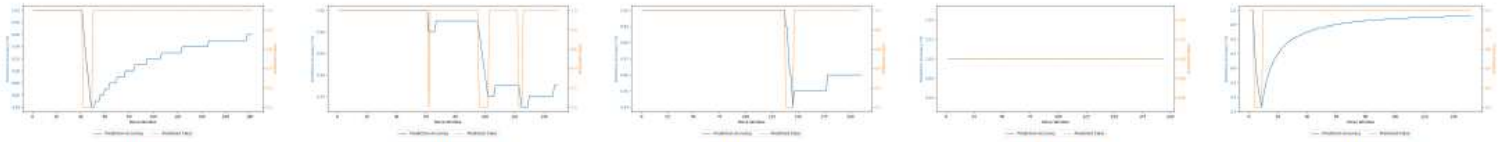
Fig. B9: Individual classification accuracy of all participants in Testset 9 where the orange line indicates the predicted class and the blue line indicates the classification accuracy

Figure B9 shows the classification accuracies obtained for the participants in Testset 9. Among the 'Solvers', one 'Expert' had 100% classification accuracy. In case of other 'Solvers', the classification accuracy ranges between 80% and 98%. Among the 'Non-solvers', all 'Quitters' and two 'Triers' had 100% classification accuracy.



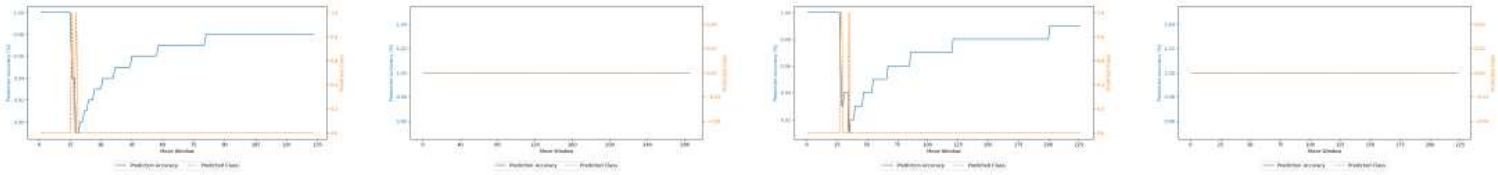
(a) Expert 1 (b) Expert 2 (c) Expert 3 (d) Expert 4

(a) Individual classification accuracy plot of all Experts



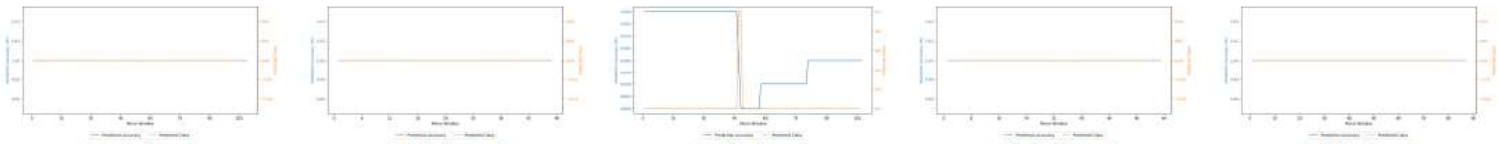
(a) Player 1 (b) Player 2 (c) Player 3 (d) Player 4 (e) Player 5

(b) Individual classification accuracy of all Persistent players



(a) Trier 1 (b) Trier 2 (c) Trier 3 (d) Trier 4

(c) Individual classification accuracy of all Triers



(a) Quitter 1 (b) Quitter 2 (c) Quitter 3 (d) Quitter 4 (e) Quitter 5

(d) Individual classification accuracy of all Quitters

Fig. B10: Individual classification accuracy of all participants in Testset 10 where the orange line indicates the predicted class and the blue line indicates the classification accuracy

Among all the participants in Testset 10, three 'Experts', one 'Persistent' player, two 'Triers', and four 'Quitters' obtained 100%. For the rest of the participants, the saddle point was obtained at an early stage, but their gaze behavior fluctuated between the gameplay, hence the classifier did not obtain a 100% accuracy.

B.2 Individual Remaining Moves

This section shows all the figures obtained by plotting the predicted remaining moves for all testsets prepared by a data preparation window size of 8 moves. The blue line indicates the moves predicted by Random-Forest regressor trained on all participants, the green line indicates the moves predicted by Random-Forest regressor trained on 'Solvers', and the orange line indicates the moves predicted by Random-Forest model trained on only 'Non-solvers'. Figure B11 shows the predicted number of moves for all

participants in Testset 1. It is seen that the 'Experts' and 'Quitters' had less erratic graphs, and 'Expert 2' and 'Expert 4' had fewer actual moves and the predicted moves graph is very smooth, and the model that was trained on only solvers produced a straight line for these two participants. 'Persistent' players 1, player 2 and player 4 had more than 250 actual moves, among which two players had a very erratic graph, and another player who had taken around 180 actual moves to solve the puzzle also had an erratic graph. Among the 'Non-solvers', all the 'Quitters' in this group had a higher number of predicted moves when Random-Forest model was trained with 'Solvers'. From the graph, a clearly visible difference can be found between the prediction of the models trained on only 'Solver' and the model trained on only 'Non-solver'. In the case of the 'Triers', it can be seen that there was a difference between the predicted moves of the previously mentioned models, but in the end, both those models gave similar predicted moves.

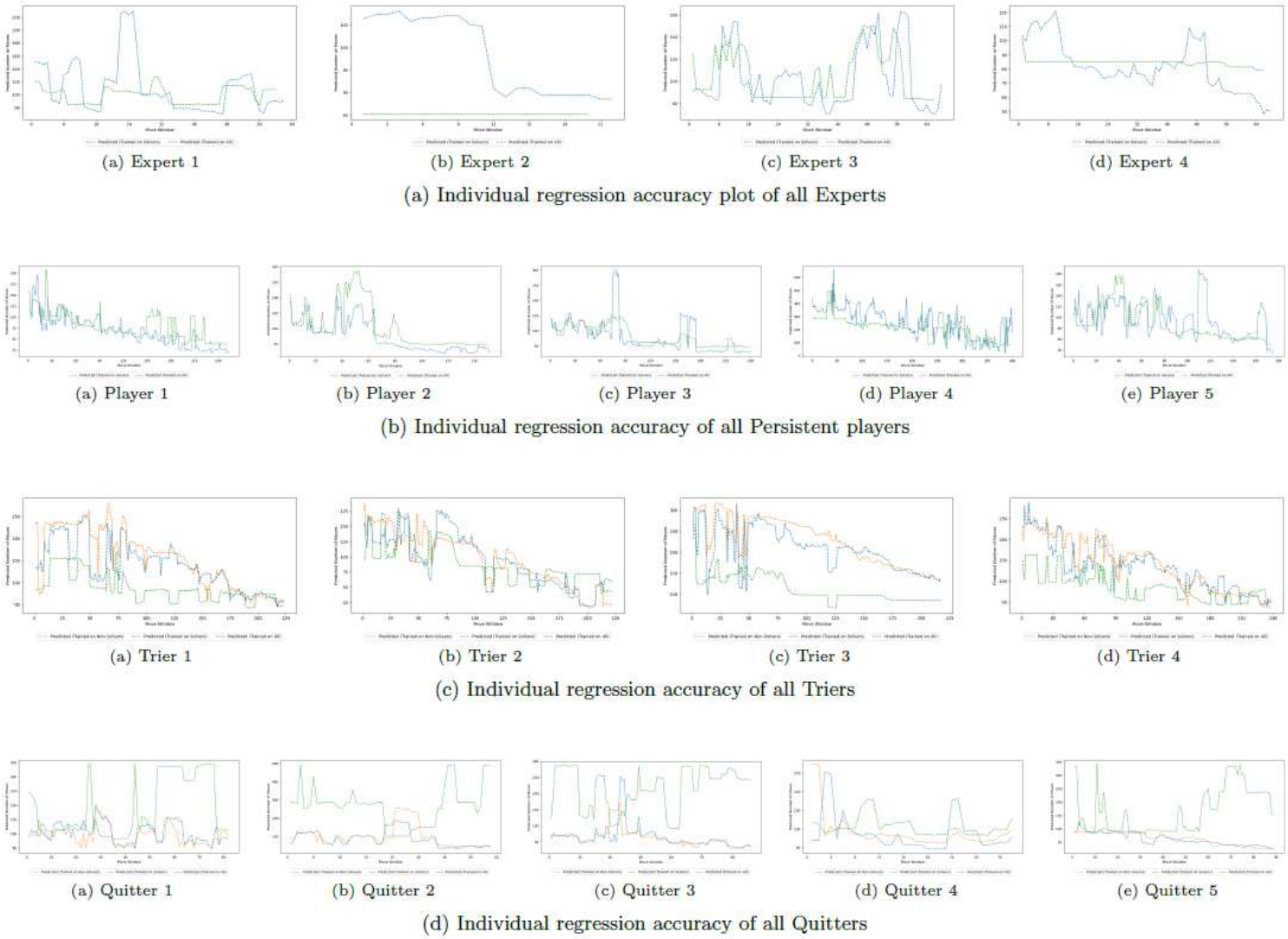
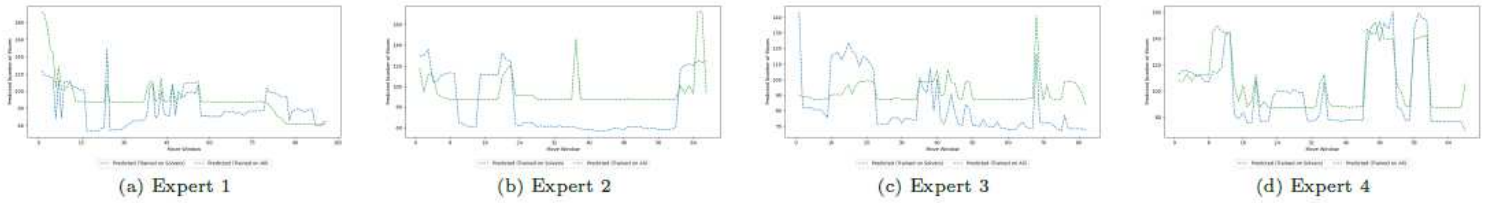


Fig. B11: Individual predicted moves of all participants in Testset 1

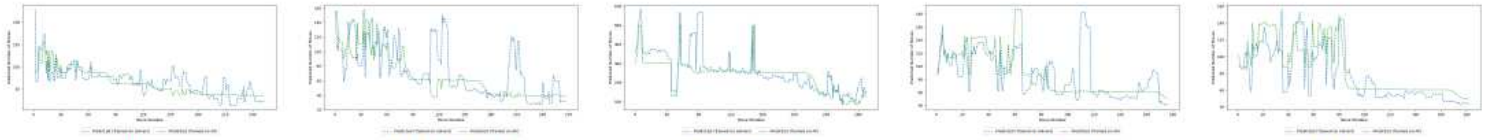
Figure B12 shows the graphs obtained by plotting the predicted remaining moves of the participants in Testset 2. All the 'Experts' in this group had similar graphs; in this case of three 'Experts', Random-Forest regressor, trained on 'Solvers' gave a

higher predicted moves than the actual predicted moves whereas Random-forest regressor trained on all participants predicted the moves better than the other model, this is because the model trained on only 'Solvers' had an overfitting issue, whereas the other model learned a more general pattern. All the 'Persistent' players in this group had very erratic graphs. In the case of the 'Quitters' in this group, the difference between the prediction of the model trained on only 'Solvers' and the prediction of the model trained on only 'Non-solvers' or all participants is visible, as in the previous set. Here, in the case of two 'Triers', the predicted remaining moves increased towards the end of the gameplay and for the other two participants, it decreased. Interestingly, for these two 'Triers', the moves predicted by all three models trained with either 'Solvers', 'Non-solvers' or all, decreased towards the end, which indicates that, though these participants left the game, they were taking the right moves at this time.



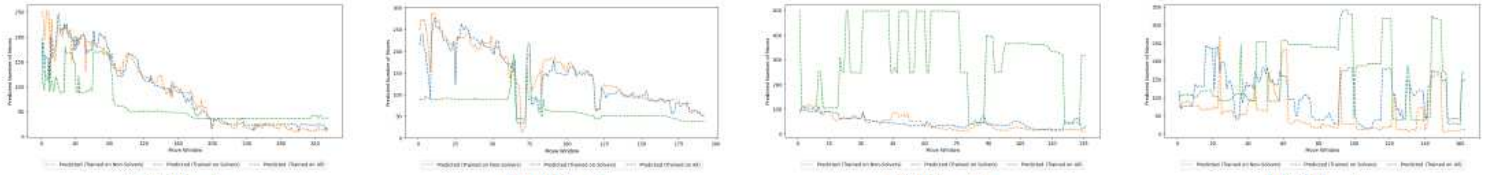
(a) Expert 1 (b) Expert 2 (c) Expert 3 (d) Expert 4

(a) Individual regression accuracy plot of all Experts



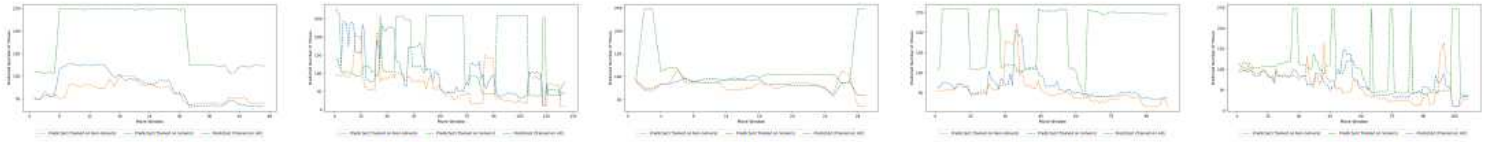
(a) Player 1 (b) Player 2 (c) Player 3 (d) Player 4 (e) Player 5

(b) Individual regression accuracy of all Persistent players



(a) Trier 1 (b) Trier 2 (c) Trier 3 (d) Trier 4

(c) Individual regression accuracy of all Triers



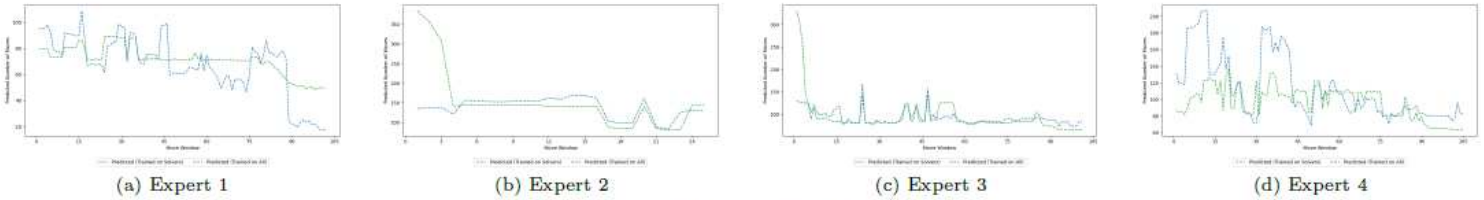
(a) Quitter 1 (b) Quitter 2 (c) Quitter 3 (d) Quitter 4 (e) Quitter 5

(d) Individual regression accuracy of all Quitters

Fig. B12: Individual predicted moves of all participants in Testset 2

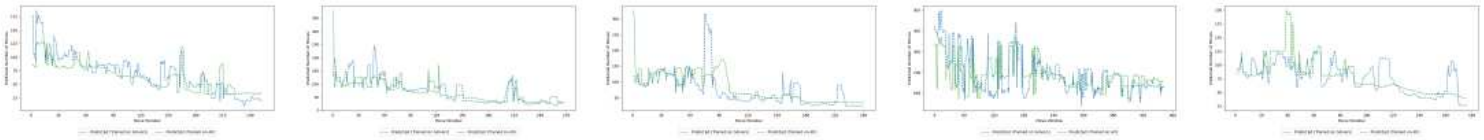
Figure B13 shows all the graphs obtained for different participants in Testset 3. Among the 'Solvers' one 'Persistent' player had an extremely erratic graph, and that participant originally took nearly 480 moves to solve the puzzle and one 'Expert' had a graph

that did not increase or decrease that much; it was nearly a straight line with some fluctuations. As the participant had a very small number of actual moves before the model could identify the gaze behavior, the game was finished, hence the model could not predict properly. In case of the 'Non-solvers', all three models gave a noisy graph for two 'Triers' and the rest of the 'Non-solvers' obtained a comparatively smoother graph, where Random-Forest model trained on 'Solvers' showed an opposite trend than the models trained on 'Non-solvers' and all categories.



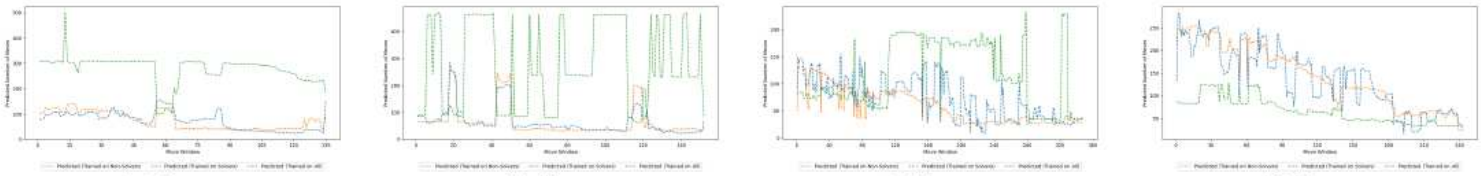
(a) Expert 1 (b) Expert 2 (c) Expert 3 (d) Expert 4

(a) Individual regression accuracy plot of all Experts



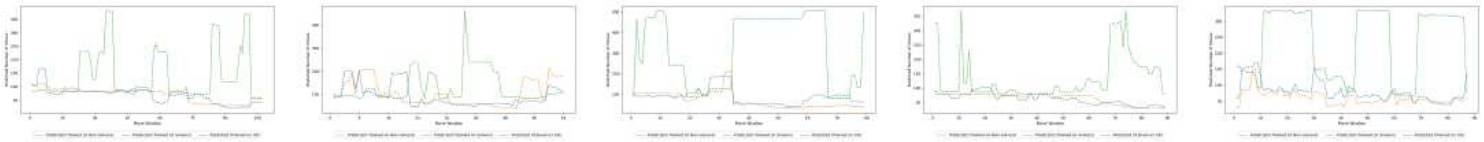
(a) Player 1 (b) Player 2 (c) Player 3 (d) Player 4 (e) Player 5

(b) Individual regression accuracy of all Persistent players



(a) Trier 1 (b) Trier 2 (c) Trier 3 (d) Trier 4

(c) Individual regression accuracy of all Triers

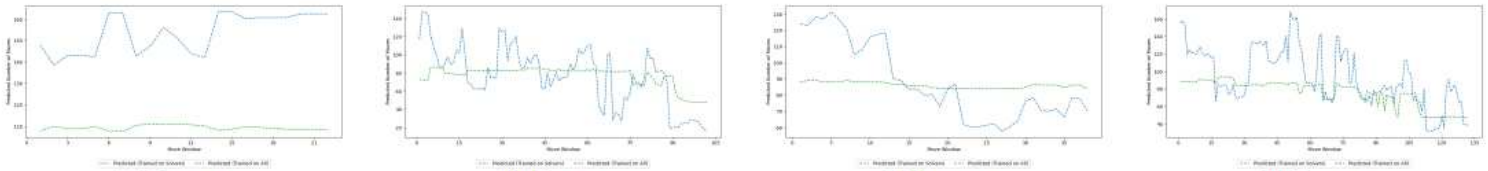


(a) Quitter 1 (b) Quitter 2 (c) Quitter 3 (d) Quitter 4 (e) Quitter 5

(d) Individual regression accuracy of all Quitters

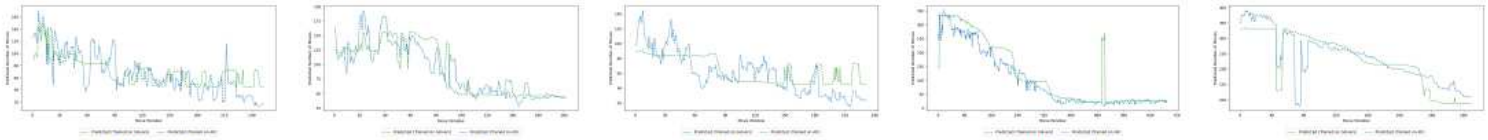
Fig. B13: Individual predicted moves of all participants in Testset 3

Figure B14 shows that all 'Persistent' players and all 'Triers' in Testset 4 obtained a noisy graph. One of the 'Quitter' in this set had a noisy graph. It is noticeable that, for those participants who had a noisy graph, the remaining moves predicted by the model trained on 'Solvers' are decreasing towards the end of the game. Two 'Experts' in this group had a very smooth curve, as the model could not identify the behavior. Other 'Experts' and 'Quitters' other than the previously mentioned one had an expected behavior.



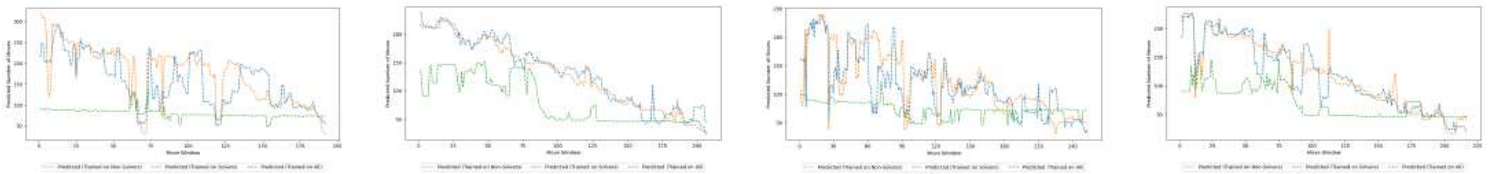
(a) Expert 1 (b) Expert 2 (c) Expert 3 (d) Expert 4

(a) Individual regression accuracy plot of all Experts



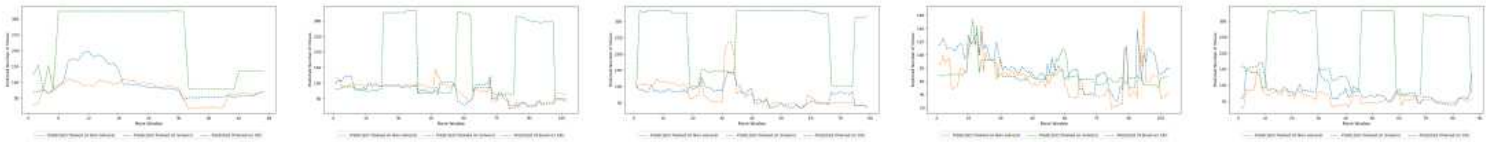
(a) Player 1 (b) Player 2 (c) Player 3 (d) Player 4 (e) Player 5

(b) Individual regression accuracy of all Persistent players



(a) Trier 1 (b) Trier 2 (c) Trier 3 (d) Trier 4

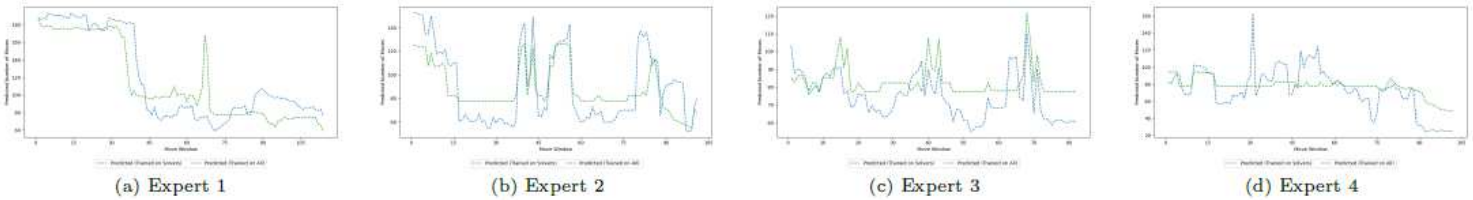
(c) Individual regression accuracy of all Triers



(a) Quitter 1 (b) Quitter 2 (c) Quitter 3 (d) Quitter 4 (e) Quitter 5

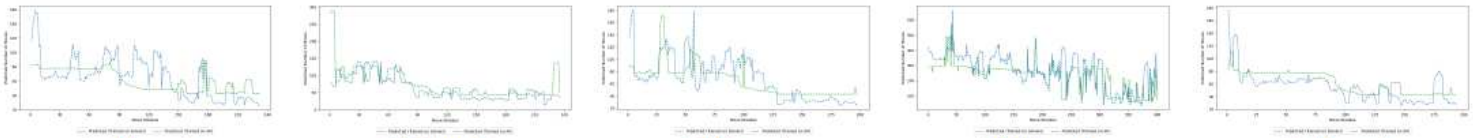
(d) Individual regression accuracy of all Quitters

Fig. B14: Individual predicted moves of all participants in Testset 4



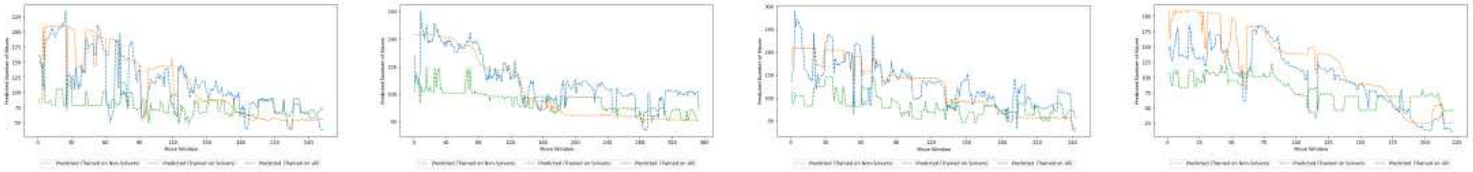
(a) Expert 1 (b) Expert 2 (c) Expert 3 (d) Expert 4

(a) Individual regression accuracy plot of all Experts



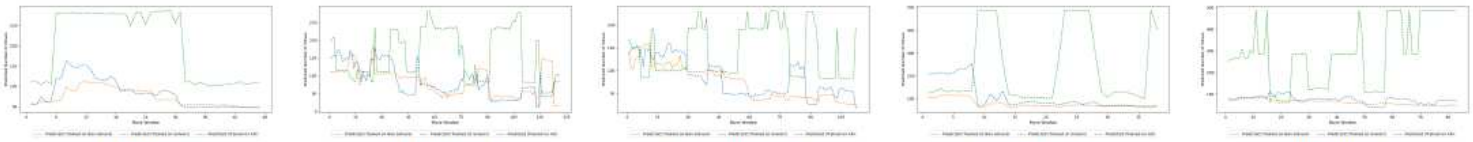
(a) Player 1 (b) Player 2 (c) Player 3 (d) Player 4 (e) Player 5

(b) Individual regression accuracy of all Persistent players



(a) Trier 1 (b) Trier 2 (c) Trier 3 (d) Trier 4

(c) Individual regression accuracy of all Triers



(a) Quitter 1 (b) Quitter 2 (c) Quitter 3 (d) Quitter 4 (e) Quitter 5

(d) Individual regression accuracy of all Quitters

Fig. B15: Individual predicted moves of all participants in Testset 5

Figure B15 has the plots obtained with the predicted remaining moves for the participants in Testset 5. The resulting graph obtained with this set became noisier as the original remaining moves increased. The fourth player in the 'Persistent' group had the noisiest curve, and the original moves that the participant had to complete the game was around 400. The 'Experts' in this group took 80-100 moves to complete the game, and for those participants, the prediction fluctuated a bit but not as much as it did for the persistent players. In case of the 'Non-solvers' in this group, the graphs obtained for the 'Triers' are noisier than the 'Quitters like the previous testsets.

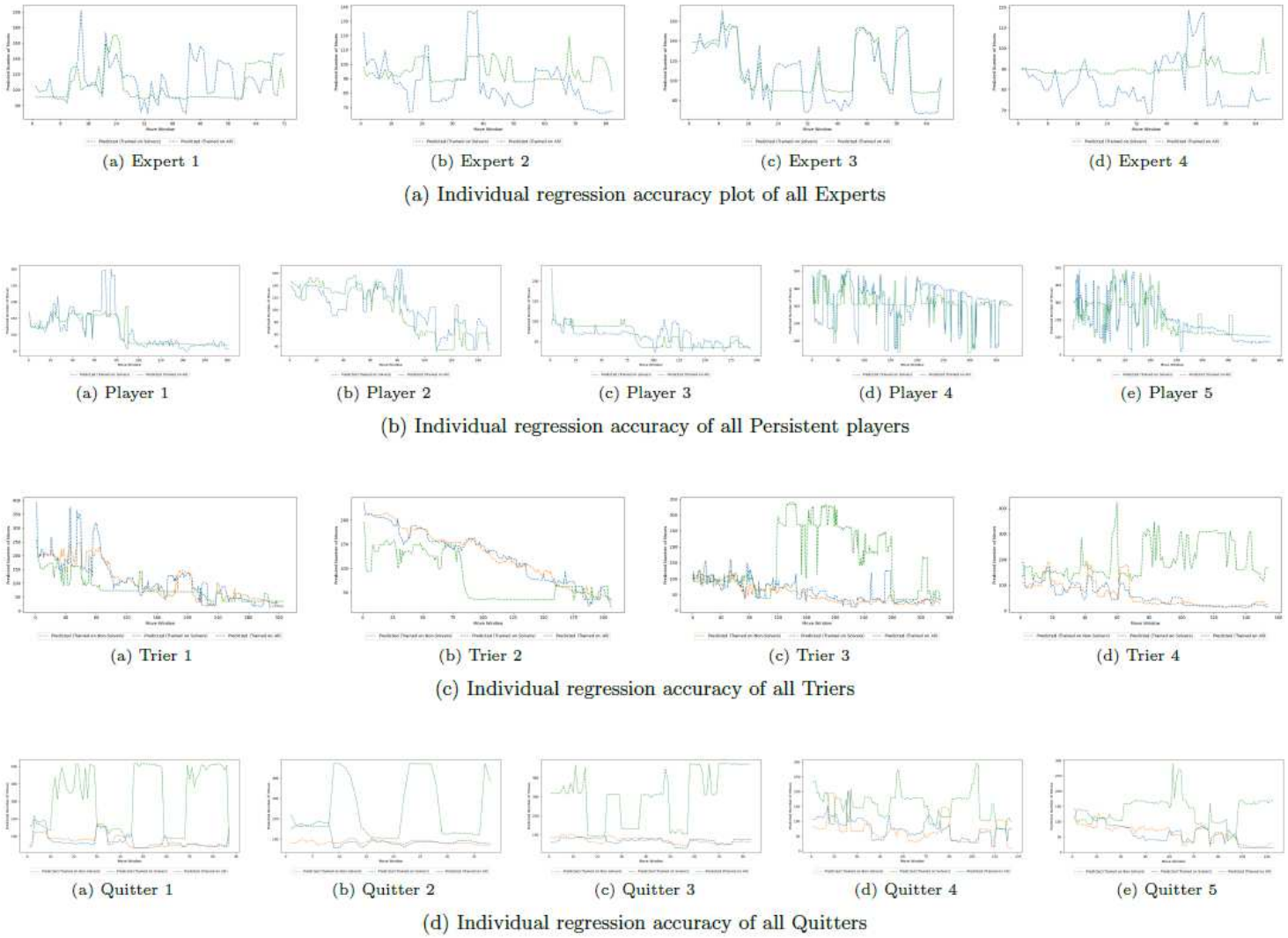
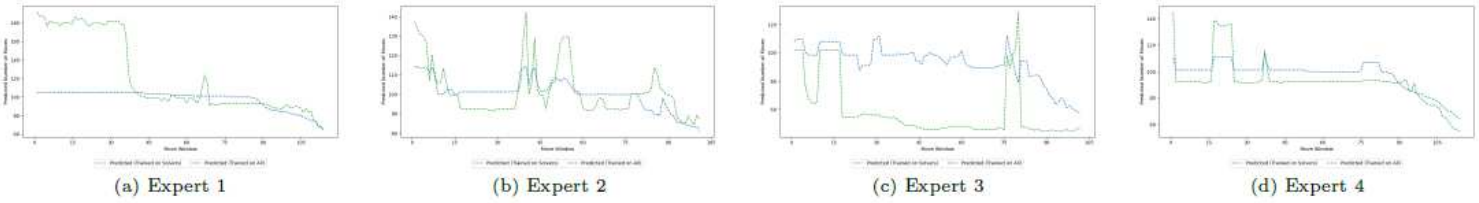


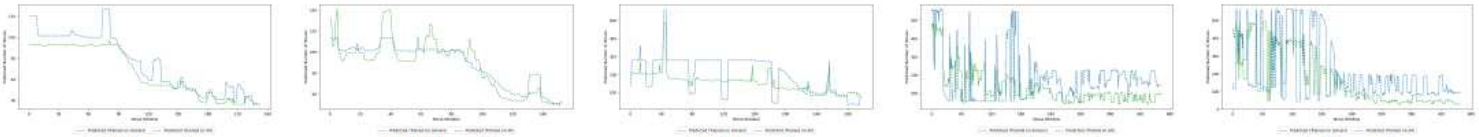
Fig. B16: Individual predicted moves of all participants in Testset 6

Figure B16 has all graphs obtained with the predicted remaining moves for Testset 6. Among the 'Solvers', the fourth and fifth 'Persistent' players originally took between 380-460 moves and had the noisiest graph. The 'Experts' in this group originally took between 70-80 moves, and the other 'Persistent' players originally took fewer than 200 moves. The amount of fluctuations in the prediction for these participants was similar. Among the 'Non-solvers', 'Quitter 2' originally took around 40 moves; hence, Random-Forest regressor model could not understand the pattern of the player. For the rest of the 'Quitters' an expected result was obtained, i.e., the graphs are smoother and the remaining moves predicted by the model trained on only 'Solvers' were increased towards the end. In the case of the 'Triers', for Trier 1 and Trier 2, the predicted moves were continuously decreasing, but in case of the other two 'Triers', the model trained on only 'solvers' gave different behavior. For one of them, the predicted moves increased suddenly in the middle of the gameplay and then decreased again, and in the case of the other 'Trier', the predicted moves started to increase and did not drop after that.



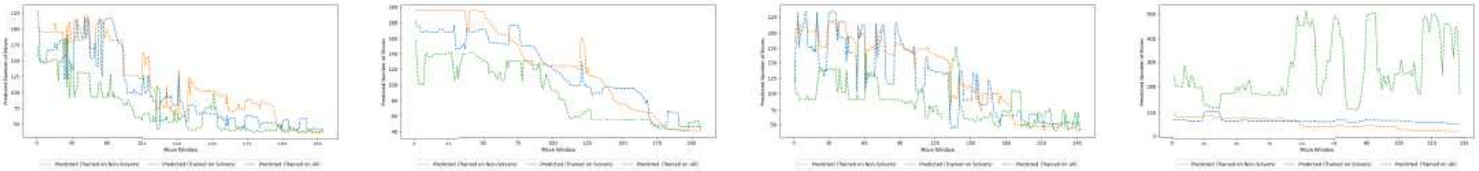
(a) Expert 1 (b) Expert 2 (c) Expert 3 (d) Expert 4

(a) Individual regression accuracy plot of all Experts



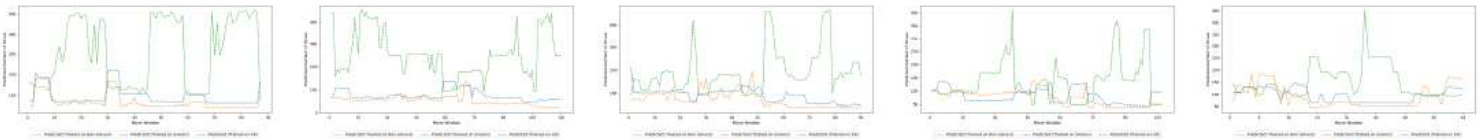
(a) Player 1 (b) Player 2 (c) Player 3 (d) Player 4 (e) Player 5

(b) Individual regression accuracy of all Persistent players



(a) Trier 1 (b) Trier 2 (c) Trier 3 (d) Trier 4

(c) Individual regression accuracy of all Triers

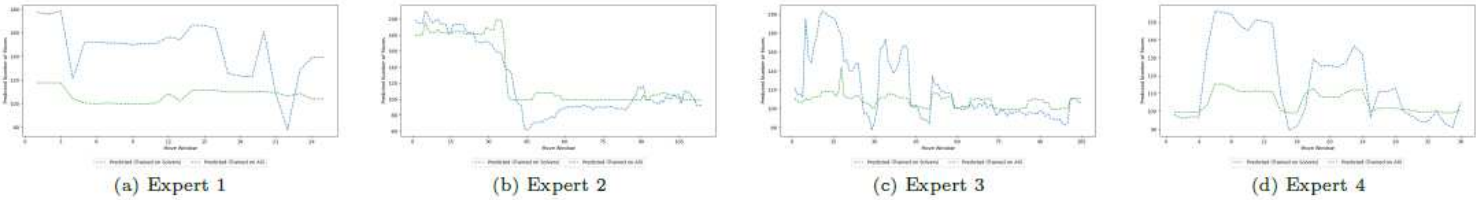


(a) Quitter 1 (b) Quitter 2 (c) Quitter 3 (d) Quitter 4 (e) Quitter 5

(d) Individual regression accuracy of all Quitters

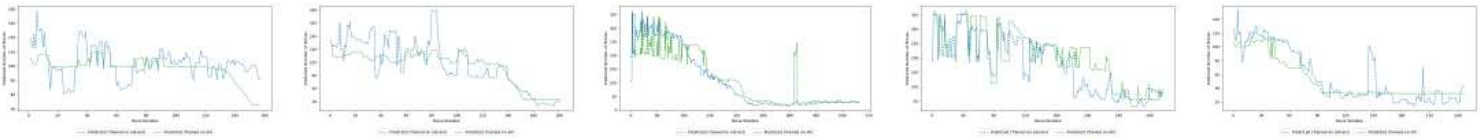
Fig. B17: Individual predicted moves of all participants in Testset 7

Figure B17 shows the predicted remaining moves plots obtained for different participants in Testset 7. Among the 'Solvers', two 'Persistent' players had more than 450 moves, two players took around 250 moves, one had taken around 140 moves and all the 'Experts ' had around 105 original moves. Figure B17 shows that all participants except two 'Persistent' players, who took over 450 moves to solve the puzzle, had a similar amount of noise in their graph. It indicates that unless the difference between the original moves is huge, the amount of noise that will be present in the predicted moves graph is not dependent on the original moves. In case of the 'Non-solvers', two 'Triers' had more than 250 original moves, one had around 200 moves, and one had around 135 moves, and the original moves for the 'Quitters' ranged between 60-105. It can be seen in the figure that the participants who had taken fewer than 140 moves had a huge difference between the prediction of the model trained on 'Solver' and the prediction of the models trained on either 'Non-solver' or all participants.



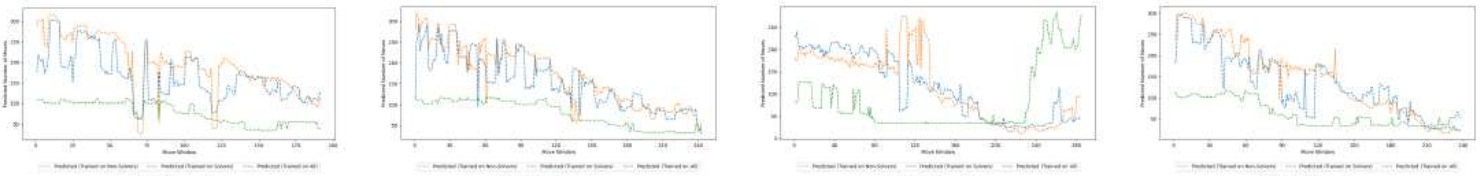
(a) Expert 1 (b) Expert 2 (c) Expert 3 (d) Expert 4

(a) Individual regression accuracy plot of all Experts



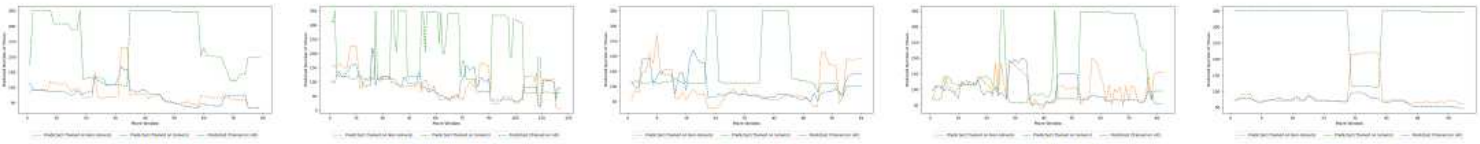
(a) Player 1 (b) Player 2 (c) Player 3 (d) Player 4 (e) Player 5

(b) Individual regression accuracy of all Persistent players



(a) Trier 1 (b) Trier 2 (c) Trier 3 (d) Trier 4

(c) Individual regression accuracy of all Triers



(a) Quitter 1 (b) Quitter 2 (c) Quitter 3 (d) Quitter 4 (e) Quitter 5

(d) Individual regression accuracy of all Quitters

Fig. B18: Individual predicted moves of all participants in Testset 8

Figure B18 has all the graphs obtained for Testset 8 after plotting the predicted remaining moves through actual gameplay. In the 'Solver' group, two 'Persistent' players had more than 250 moves, one player took above 650 moves, and others took around 160-180 moves; all the participants had a noisier graph compared to the 'Experts' and the players with the highest number of moves had more noise in their graph. In the 'Expert' group, two participants had fewer than 40 original moves and in their case, the model initially gave a prediction that they would take more than 100-150 moves, but the model could not settle down in its prediction, as the original number of moves is very low. In case of the 'Non-solvers', the regressor produced the expected result, i.e., the 'Triers' had a noisier graph and the 'Quitters' showed a difference between the prediction of two models, one trained on 'Solvers' only, and the other trained on 'Non-solvers' only or all categories. Here, in the case of 'Trier 3', the predicted moves by Random-Forest regressor trained on 'Solver' increased towards the end of the gameplay.

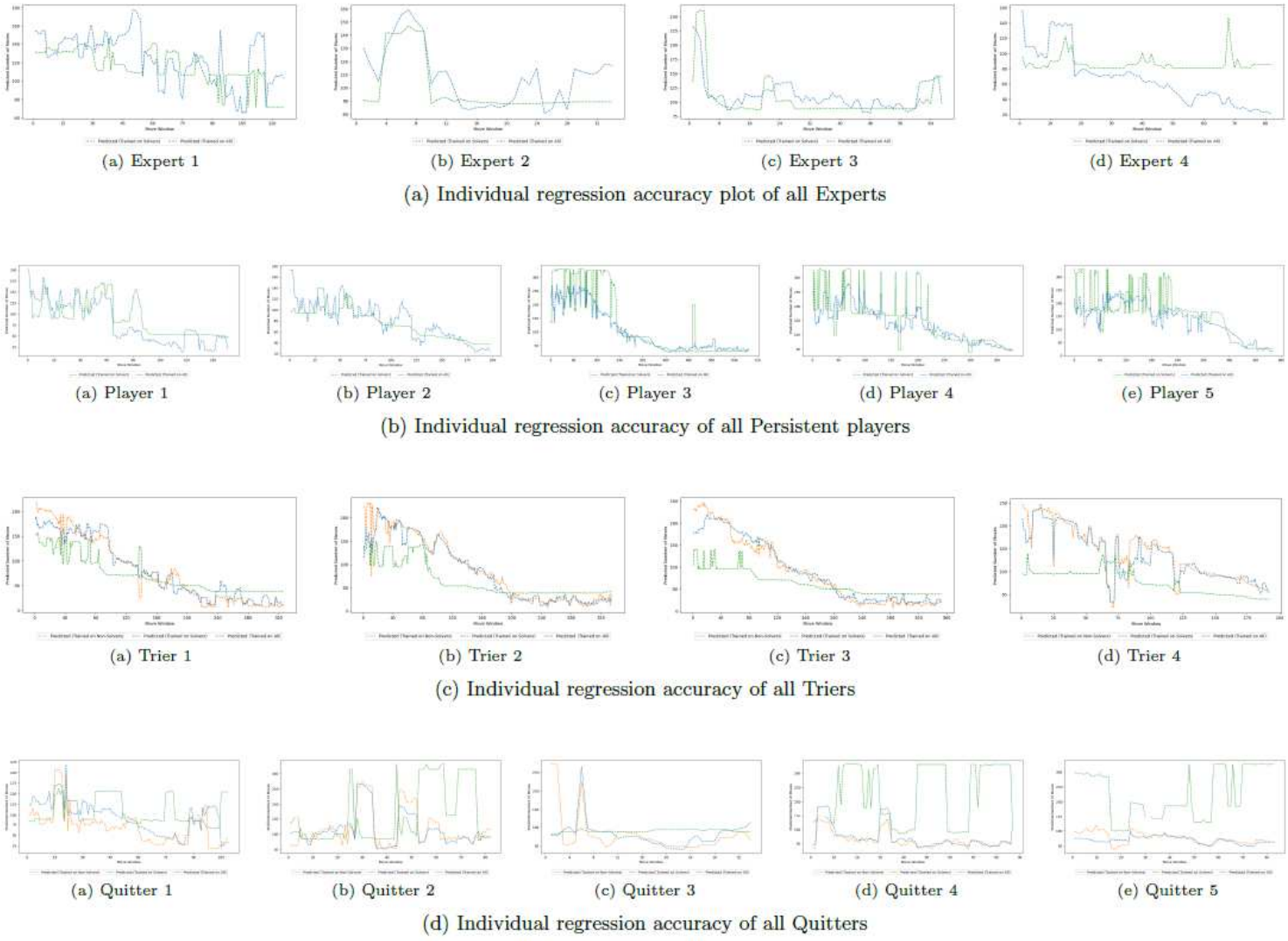
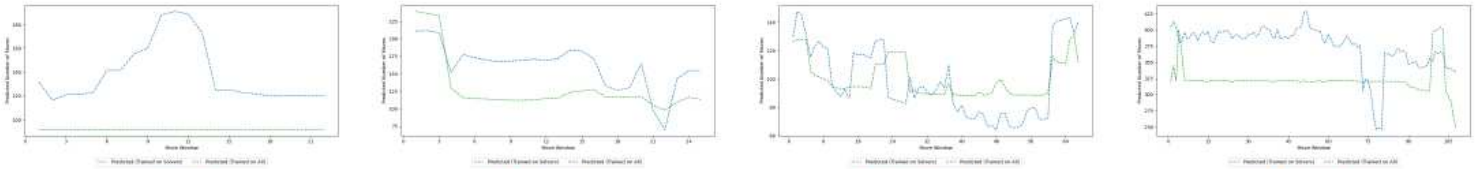


Fig. B19: Individual predicted moves of all participants in Testset 9

Figure B19 has all the graphs obtained for Testset 9 after plotting the predicted remaining moves through actual gameplay. This group had one 'Expert' with fewer than 40 moves for whom the regressor could not predict the remaining moves properly, and the original moves of the other three 'Experts' ranged from 60 to 120; these three 'Experts' had a comparatively smoother graph than the 'Persistent' players in this group. In the 'Persistent' group, three players had more than 350 original moves, who had noisier graphs, and two participants had around 140-200 moves, who obtained graphs smoother than the other three 'Persistent' players. Among the 'Non-solvers', the original moves of the 'Triers' ranged between 200-360, and the original moves of the 'Quitters' ranged between 30-110. Here we found an expected behavior, i.e., the 'Triers' had a noisier graph and the 'Quitters' showed a difference between the prediction of two models, one trained on 'Solvers' only, and the other trained on 'Non-solvers' only or all categories.

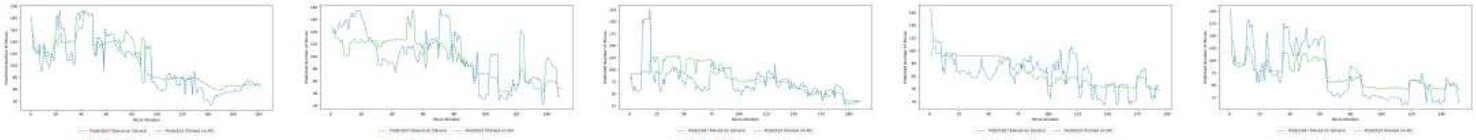
Figure B20 has all the graphs obtained for Testset 10, after plotting the predicted remaining moves through actual gameplay. In this group, two 'Experts' completed the game within 25 moves, and as Random-Forest regressor needs some time to understand the gameplay behavior, it failed to predict their behavior properly; it predicted that those participants would take around 130 and

200 moves to finish. The other 'Experts' in this group took around 70-120 moves. In those cases, the regressor produced a smooth curve and predicted the behavior properly. In case of the 'Non-solvers', the result of this set was similar to the result of Testset 7, Testset 10 had one 'Trier' with around 135 moves and three 'Triers' with more than 200 moves, in the case of the 'Triers ' with a higher number of moves, the graph had more noise. For the 'Quitters', we found an expected behavior, i.e., the 'Triers' had a noisier graph, and these 'Quitters' had a smoother graph, and they show a difference between the prediction of two models, one trained on 'Solvers' only, and the other trained on 'Non-solvers' only or all categories.



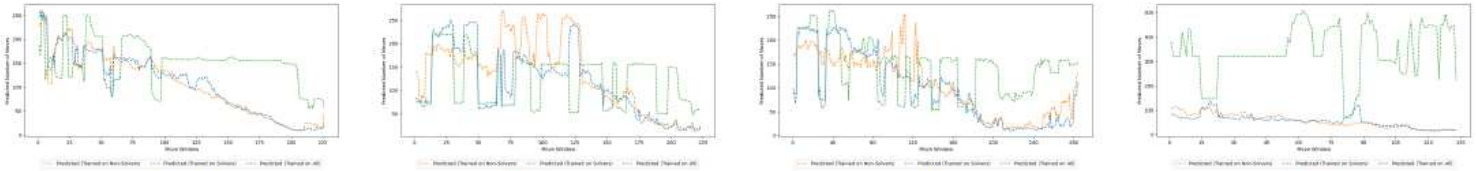
(a) Expert 1 (b) Expert 2 (c) Expert 3 (d) Expert 4

(a) Individual regression accuracy plot of all Experts



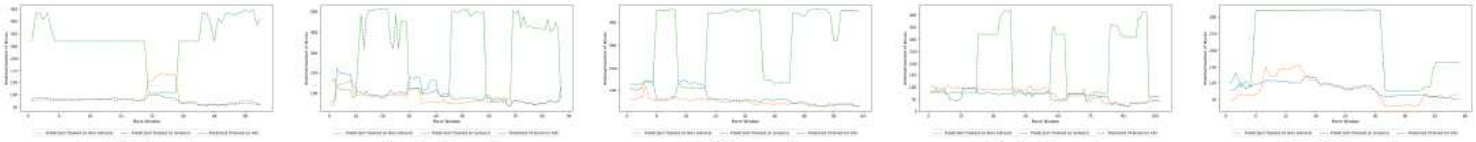
(a) Player 1 (b) Player 2 (c) Player 3 (d) Player 4 (e) Player 5

(b) Individual regression accuracy of all Persistent players



(a) Trier 1 (b) Trier 2 (c) Trier 3 (d) Trier 4

(c) Individual regression accuracy of all Triers



(a) Quitter 1 (b) Quitter 2 (c) Quitter 3 (d) Quitter 4 (e) Quitter 5

(d) Individual regression accuracy of all Quitters

Fig. B20: Individual predicted moves of all participants in Testset 10

Appendix C Additional Tables

C.1 Difference between Experts and Persistent Group

This subsection contains the standard deviation, coefficient of variation, and signal-to-noise ratio values of all the 'Solvers' in all test sets, across all data preparation windows except window size 8.

The signal-to-noise ratio is infinity for the data preparation window size 30 as shown in Table C8. These test sets had participants who completed the game in 30 moves; hence, the data preparation window 30 produces only one data sample for these participants. The Signal-to-Noise ratio is calculated by dividing the mean by the variance, and in these cases the variance is zero, hence the Signal-to-Noise ratio is infinity. All testsets obtained will different data preparation window shows that the standard deviation value and coefficient of variation of the 'Experts' are less than that of the 'Persistent players' and the signal-to-noise ratio of the 'Experts' are higher than that of the 'Persistent players', this indicates that the graph obtained for the 'Experts' are less noisier than the 'Persistent players'.

Table C2: Mean Standard deviation, Mean Coefficient variation and Mean Signal-to-noise ratio obtained with data preparation window size 6 for all solvers in different test sets

Set	Expert Std	PP Std	Expert CV	PP CV	Expert SNR	PP SNR
Set 1	21.610	58.122	0.206	0.534	6.698	1.954
Set 2	23.817	45.956	0.282	0.448	3.960	2.291
Set 3	25.926	48.604	0.274	0.544	3.974	1.891
Set 4	24.381	57.088	0.283	0.548	4.920	2.150
Set 5	27.689	47.219	0.301	0.473	3.538	2.150
Set 6	20.860	66.922	0.197	0.425	6.236	2.486
Set 7	13.608	87.329	0.148	0.515	7.802	2.385
Set 8	28.485	63.006	0.228	0.569	5.050	2.353
Set 9	35.891	66.106	0.362	0.598	3.113	1.753
Set 10	23.507	36.123	0.156	0.458	7.997	2.258

Table C3: Mean Standard deviation, Mean Coefficient variation and Mean Signal-to-noise ratio obtained with data preparation window size 7 for all solvers in different test sets

Set	Expert Std	PP Std	Expert CV	PP CV	Expert SNR	PP SNR
Set 1	21.676	59.843	0.211	0.554	6.641	1.911
Set 2	23.533	50.685	0.281	0.475	4.102	2.137
Set 3	26.712	50.459	0.272	0.545	3.906	1.897
Set 4	22.788	57.532	0.264	0.566	7.554	1.994
Set 5	27.884	47.660	0.305	0.477	3.790	2.170
Set 6	21.751	68.726	0.209	0.431	5.443	2.458
Set 7	11.776	91.046	0.123	0.530	9.271	2.358
Set 8	30.626	62.138	0.244	0.571	4.425	2.169
Set 9	30.374	59.965	0.303	0.548	3.486	1.941
Set 10	21.140	36.782	0.133	0.463	10.015	2.282

Table C4: Mean Standard deviation, Mean Coefficient variation and Mean Signal-to-noise ratio obtained with data preparation window size 10 for all solvers in different test sets

Set	Expert Std	PP Std	Expert CV	PP CV	Expert SNR	PP SNR
Set 1	17.798	52.202	0.177	0.504	12.677	2.096
Set 2	27.299	56.297	0.327	0.480	3.279	2.163
Set 3	28.00	56.221	0.276	0.570	4.082	1.804
Set 4	20.134	59.021	0.234	0.567	7.927	2.014
Set 5	32.470	45.235	0.352	0.516	3.231	2.007
Set 6	24.600	67.938	0.232	0.458	6.163	2.246
Set 7	24.243	85.082	0.258	0.606	4.564	2.093
Set 8	36.611	65.215	0.295	0.591	4.088	2.102
Set 9	27.901	55.281	0.291	0.503	3.733	2.101
Set 10	38.255	39.208	0.213	0.477	5.093	2.174

Table C5: Mean Standard deviation, Mean Coefficient variation and Mean Signal-to-noise ratio obtained with data preparation window size 12 for all solvers in different test sets

Set	Expert Std	PP Std	Expert CV	PP CV	Expert SNR	PP SNR
Set 1	15.151	46.484	0.163	0.439	135	2.293
Set 2	19.313	44.204	0.206	0.409	5.170	2.453
Set 3	24.175	52.134	0.278	0.507	4.055	2.037
Set 4	22.891	63.398	0.243	0.595	6.415	1.882
Set 5	40.556	47.002	0.406	0.530	2.633	1.989
Set 6	25.298	71.856	0.246	0.509	4.330	2.112
Set 7	28.661	85.789	0.320	0.650	3.309	1.845
Set 8	36.154	58.471	0.306	0.543	4.153	2.123
Set 9	32.488	46.143	0.331	0.447	3.595	2.326
Set 10	28.291	39.769	0.192	0.455	5.908	2.255

Table C6: Mean Standard deviation, Mean Coefficient variation and Mean Signal-to-noise ratio obtained with data preparation window size 15 for all solvers in different test sets

Set	Expert Std	PP Std	Expert CV	PP CV	Expert SNR	PP SNR
Set 1	21.621	48.060	0.198	0.466	6.966	2.516
Set 2	32.586	53.093	0.270	0.471	4.603	2.202
Set 3	26.445	44.698	0.253	0.431	4.805	2.365
Set 4	20.9	63.363	0.221	0.579	6.748	1.994
Set 5	35.663	42.448	0.369	0.460	2.822	2.206
Set 6	23.579	69.942	0.224	0.490	4.884	2.167
Set 7	26.751	74.941	0.275	0.596	4.140	1.953
Set 8	38.677	52.123	0.337	0.497	3.658	2.512
Set 9	30.176	58.499	0.282	0.510	4.703	2.010
Set 10	36.073	40.680	0.201	0.439	6.966	2.349

Table C7: Mean Standard deviation, Mean Coefficient variation and Mean Signal-to-noise ratio obtained with data preparation window size 20 for all solvers in different test sets

Set	Expert Std	PP Std	Expert CV	PP CV	Expert SNR	PP SNR
Set 1	13.135	50.038	0.124	0.434	9.587	2.617
Set 2	40.915	51.168	0.317	0.448	4.558	2.489
Set 3	21.398	53.926	0.204	0.515	6.554	2.091
Set 4	20.664	67.400	0.221	0.662	5.244	1.669
Set 5	20.506	48.782	0.230	0.514	4.648	1.989
Set 6	20.579	65.216	0.183	0.513	5.672	2.081
Set 7	23.392	64.765	0.256	0.565	4.146	2.242
Set 8	31.635	41.709	0.263	0.439	4.417	2.901
Set 9	14.387	43.996	0.151	0.450	8.289	2.287
Set 10	36.383	42.881	0.203	0.423	13.262	2.492

Table C8: Mean Standard deviation, Mean Coefficient variation and Mean Signal-to-noise ratio obtained with data preparation window size 30 for all solvers in different test sets

Set	Expert Std	PP Std	Expert CV	PP CV	Expert SNR	PP SNR
Set 1	20.339	58.162	0.238	0.452	inf	2.405
Set 2	22.095	43.546	0.261	0.401	4.590	2.526
Set 3	13.640	54.667	0.152	0.474	24.329	2.389
Set 4	22.459	48.128	0.198	0.498	inf	2.063
Set 5	20.547	54.192	0.267	0.533	4.307	1.917
Set 6	20.153	60.800	0.212	0.449	5.233	2.604
Set 7	11.233	63.470	0.143	0.517	13.097	2.986
Set 8	26.909	41.566	0.194	0.421	6.407	2.537
Set 9	18.170	46.918	0.201	0.448	5.780	2.259
Set 10	38.882	44.667	0.201	0.498	inf	2.252

C.2 Difference between Triers and Quitters

This subsection contained the standard deviation, coefficient of variation and the Signal-to-noise ratio values of all the Non-solvers in all test sets across all data preparation windows except window size 8. All tables from C9 to C15 shows that except for Testset 1 prepared with data preparation window 12 and 15, all testsets prepared with different data preparation windows have higher standard deviation and coefficient of variation values for the 'Triers' than the 'Quitters' and have a lower signal-to-noise ratio for 'Triers' than the 'Quitters' which indicates that the graphs obtained with Random-Forest regressor results, have higher amount of noise than the 'Quitters'.

Table C9: Mean Standard deviation, Mean Coefficient variation and Mean Signal-to-noise ratio obtained with data preparation window size 6 for all non-solvers in different test sets

Set	Trier Std	Quitter Std	Trier CV	Quitter CV	Trier SNR	Quitter SNR
Set 1	54.196	37.311	0.379	0.436	2.802	2.374
Set 2	52.747	26.598	0.614	0.381	1.726	3.122
Set 3	48.297	37.063	0.615	0.487	1.703	2.189
Set 4	57.596	36.565	0.436	0.466	2.315	2.194
Set 5	54.834	33.253	0.462	0.409	2.181	2.548
Set 6	51.648	33.854	0.593	0.468	1.737	2.310
Set 7	49.944	26.792	0.454	0.325	2.245	3.525
Set 8	69.810	30.737	0.505	0.390	2.067	2.752
Set 9	64.834	34.974	0.672	0.426	1.631	2.661
Set 10	60.940	30.278	0.598	0.401	1.708	2.752

Table C10: Mean Standard deviation, Mean Coefficient variation and Mean Signal-to-noise ratio obtained with data preparation window size 7 for all non-solvers in different test sets

Set	Trier Std	Quitter Std	Trier CV	Quitter CV	Trier SNR	Quitter SNR
Set 1	51.206	39.402	0.369	0.453	2.912	2.288
Set 2	57.317	33.538	0.613	0.428	1.716	2.675
Set 3	49.459	38.483	0.613	0.495	1.714	2.130
Set 4	57.012	35.184	0.437	0.441	2.309	2.313
Set 5	50.535	35.502	0.436	0.415	2.312	2.784
Set 6	51.555	33.272	0.603	0.455	1.718	2.513
Set 7	42.841	32.598	0.392	0.369	3.084	2.835
Set 8	66.660	35.493	0.490	0.410	2.134	2.635
Set 9	61.679	34.240	0.651	0.406	1.661	2.798
Set 10	59.200	27.494	0.597	0.368	1.720	3.060

Table C11: Mean Standard deviation, Mean Coefficient variation and Mean Signal-to-noise ratio obtained with data preparation window size 10 for all non-solvers in different test sets

Set	Trier Std	Quitter Std	Trier CV	Quitter CV	Trier SNR	Quitter SNR
Set 1	50.715	32.236	0.398	0.385	2.761	2.746
Set 2	52.380	32.096	0.571	0.417	1.857	2.857
Set 3	51.652	36.476	0.552	0.444	1.931	2.353
Set 4	53.645	26.834	0.423	0.350	2.385	2.987
Set 5	47.392	37.404	0.432	0.404	2.376	2.940
Set 6	50.286	34.346	0.556	0.448	1.879	2.387
Set 7	51.323	36.482	0.496	0.425	2.057	2.468
Set 8	62.656	33.899	0.484	0.372	2.155	2.824
Set 9	59.053	38.829	0.649	0.431	1.643	2.557
Set 10	56.604	25.879	0.579	0.307	1.742	3.900

Table C12: Mean Standard deviation, Mean Coefficient variation and Mean Signal-to-noise ratio obtained with data preparation window size 12 for all non-solvers in different test sets

Set	Trier Std	Quitter Std	Trier CV	Quitter CV	Trier SNR	Quitter SNR
Set 1	51.062	42.968	0.391	0.470	2.755	2.147
Set 2	47.311	33.755	0.517	0.408	2.095	3.156
Set 3	47.213	34.394	0.475	0.430	2.178	2.387
Set 4	49.463	33.261	0.409	0.402	2.473	2.547
Set 5	47.177	35.880	0.438	0.395	2.310	2.572
Set 6	50.855	32.739	0.547	0.420	1.881	2.531
Set 7	49.292	53.188	0.488	0.536	2.091	2.111
Set 8	58.357	31.830	0.440	0.346	2.370	3.021
Set 9	55.827	34.829	0.618	0.379	1.727	2.781
Set 10	54.711	28.689	0.549	0.323	1.843	3.412

Table C13: Mean Standard deviation, Mean Coefficient variation and Mean Signal-to-noise ratio obtained with data preparation window size 15 for all non-solvers in different test sets

Set	Trier Std	Quitter Std	Trier CV	Quitter CV	Trier SNR	Quitter SNR
Set 1	48.259	37.119	0.391	0.405	2.684	2.595
Set 2	41.073	31.357	0.459	0.345	2.575	4.904
Set 3	50.411	33.949	0.466	0.393	2.200	2.670
Set 4	43.222	28.446	0.395	0.364	2.573	2.815
Set 5	39.885	29.925	0.370	0.306	2.753	3.344
Set 6	57.231	38.768	0.584	0.483	1.772	2.147
Set 7	44.099	51.517	0.457	0.451	2.236	2.407
Set 8	46.327	28.223	0.380	0.313	2.867	3.278
Set 9	54.510	29.781	0.611	0.361	1.737	3.187
Set 10	50.736	22.821	0.505	0.235	2.014	5.107

Table C14: Mean Standard deviation, Mean Coefficient variation and Mean Signal-to-noise ratio obtained with data preparation window size 20 for all non-solvers in different test sets

Set	Trier Std	Quitter Std	Trier CV	Quitter CV	Trier SNR	Quitter SNR
Set 1	41.784	24.552	0.365	0.284	2.793	3.753
Set 2	37.360	24.061	0.417	0.300	2.706	4.925
Set 3	44.107	38.317	0.380	0.359	2.807	2.909
Set 4	36.063	22.437	0.345	0.282	2.972	3.978
Set 5	41.546	27.233	0.408	0.305	2.482	3.301
Set 6	56.685	32.461	0.546	0.353	1.867	2.903
Set 7	55.205	38.402	0.563	0.342	1.825	3.263
Set 8	45.492	40.299	0.355	0.395	2.966	2.738
Set 9	40.623	24.944	0.495	0.303	2.240	3.462
Set 10	53.295	29.276	0.544	0.231	1.882	5.705

Table C15: Mean Standard deviation, Mean Coefficient variation and Mean Signal-to-noise ratio obtained with data preparation window size 30 for all non-solvers in different test sets

Set	Trier Std	Quitter Std	Trier CV	Quitter CV	Trier SNR	Quitter SNR
Set 1	44.365	19.907	0.417	0.229	2.418	5.051
Set 2	39.665	14.387	0.445	0.193	2.464	20.714
Set 3	34.366	29.742	0.283	0.293	3.795	3.694
Set 4	35.845	25.153	0.377	0.321	2.828	3.295
Set 5	28.779	19.899	0.320	0.239	3.381	5.137
Set 6	54.518	25.926	0.496	0.308	2.152	3.347
Set 7	28.300	37.610	0.305	0.331	4.306	3.252
Set 8	28.328	25.756	0.233	0.265	4.571	4.013
Set 9	31.409	23.538	0.422	0.260	2.865	4.937
Set 10	37.420	35.729	0.403	0.341	2.798	3.957

References

Aatrai S, Jha SK, Guha R (2023) Visual perception and performance: An eye tracking study. In: Proceedings of the 2023 Symposium on Eye Tracking Research and Applications, pp 1–5 .

Aatrai S, Gayatri SP, Guha R (2024) Learning through the eyes: Unveiling oculometrics for domain-knowledge identification. In: 2024 IEEE International Conference on Advanced Learning Technologies (ICALT), IEEE, pp 287–291 .

Ariff NI, Tasir Z, Jumaat NF (2017) Students' problem solving appraisal and behaviour in asynchronous online discussion forum. Innovative Teaching and Learning Journal (ITLJ) 1(1):1–12 .

Atkinson J, Nardini M (2008) Visuospatial and visuomotor development. In: J Reed JR (ed) "Child Neuropsychology: Concepts, Theory, and Practice". Wiley-Blackwell, chap 9, p 183–217 .

Bastiaansen M, Hagoort P (2006) Oscillatory neuronal dynamics during language comprehension. Progress in brain research 159:179–196 .

Bengio Y, Lodi A, Prouvost A (2021) Machine learning for combinatorial optimization: a methodological tour d'horizon. European Journal of Operational Research 290(2):405–421 .

Dörner D, Güss CD (2022) Human error in complex problem solving and dynamic decision making: A taxonomy of 24 errors and a theory. Computers in human behavior reports 7:100222 .

Eichmann B, Greiff S, Naumann J, et al (2020) Exploring behavioural patterns during complex problem-solving. Journal of computer assisted learning 36(6):933–956 .

Eivazi S, Bednarik R (2010) Inferring problem solving strategies using eye-tracking: system description and evaluation. In: Proceedings of the 10th Koli Calling International Conference on Computing Education Research, pp 55–61 .

Evans JSB, Ball LJ (2010) Do people reason on the wason selection task? a new look at the data of ball et al.(2003). Quarterly Journal of Experimental Psychology 63(3):434–441 .

Goel V, Dolan RJ (2004) Differential involvement of left prefrontal cortexin inductive and deductive reasoning. Cognition 93(3):B109–B121 .

Goldberg JH, Helfman JI (2010) Comparing information graphics: a critical look at eye tracking. In: Proceedings of the 3rd BELIV'10 Workshop: BEyond time and errors: novel evaLuation methods for Information Visualization, pp 71–78 .

Graf M, Danek AH, Vaci N, et al (2023) Tracing cognitive processes in insight problem solving: Using gams and change point analysis to uncover restructuring. Journal of Intelligence 11(5):86 .

Graham R (2025) 10.3 An Information Processing Perspective on Problem-Solving – Cognitive Psychology — nmoer.pressbooks.pub. https://nmoer.pressbooks.pub/cognitivepsychology/chapter/an-information-processing-perspective-on-problem-solving/, [Accessed 13-10-2025] .

Graham T, Kleshnina M, Filar JA (2023) Where do mistakes lead? a survey of games with incompetent players. Dynamic Games and Applications 13(1):231–264 .

Gray WD, Fu WT (2004) Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. Cognitive science 28(3):359–382 .

Guo Y, Pannasch S, Helmert JR, et al (2024) Ambient and focal attention during complex problem-solving: preliminary evidence from real-world eye movement data. Frontiers in Psychology 15:1217106 .

Holmqvist K, Nyström M, Andersson R, et al (2011) Eye tracking: A comprehensive guide to methods and measures. oup Oxford .

Huang OW, Cheng HN, Chan TW (2007) Number jigsaw puzzle: A mathematical puzzle game for facilitating players' problem-solving strategies. In: 2007 First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL'07), pp 130–134, https://doi.org/10.1109/DIGITEL.2007.37 .

Kahneman D, Frederick S (2000) Choices, values, and frames. American Psychologist 55(9):697–720 .

Kahou SE, Pal C, Bouthillier X, et al (2013) Combining modality specific deep neural networks for emotion recognition in video. In: Proceedings of the 15th ACM on International conference on multimodal interaction, pp 543–550 .

Kalia V, Fuesting M, Cody M (2019) Perseverance in solving sudoku: Role of grit and cognitive flexibility in problem solving. Journal of Cognitive Psychology 31(3):370–378 .

Khakifirooz E (2015) Effects of interaction method's directness on problem solving. Master's thesis, Itä-Suomen yliopisto .

Kiefer P, Giannopoulos I, Kremer D, et al (2014) Starting to get bored: An outdoor eye tracking study of tourists exploring a city panorama. In: Proceedings of the symposium on eye tracking research and applications, pp 315–318 .

Knoblich G, Ohlsson S, Raney GE (2001) An eye movement study of insight problem solving. Memory & cognition 29(7):1000–1009 .

Kocijan V, Horvat M, Majdic G (2017) Robust sex differences in jigsaw puzzle solving—are boys really better in most visuospatial tasks? Frontiers in Behavioral Neuroscience 11:194 .

Koffka K (1912) Zur grundlegung der psychologischen gestalttheorie. Zeitschrift für Psychologie 60:1–39 .

Körner C (2011) Eye movements reveal distinct search and reasoning processes in comprehension of complex graphs. Applied Cognitive Psychology 25(6):893–905 .

Körner C, Gilchrist ID (2004) Eye movements in a simple spatial reasoning task. Perception 33(4):485–494 .

Krebs C, Falkner M, Niklaus J, et al (2021) Application of eye tracking in puzzle games for adjunct cognitive markers: pilot observational study in older adults. JMIR serious games 9(1):e24151 .

Land M, Mennie N, Rusted J (1999) The roles of vision and eye movements in the control of activities of daily living. Perception 28(11):1311–1328 .

Land MF, Hayhoe M (2001) In what ways do eye movements contribute to everyday activities? Vision research 41(25-26):3559–3565 .

Landgraf S, Amado I, Brucks M, et al (2011) Inflexible information acquisition strategies mediate visuo-spatial reasoning in stabilized schizophrenia patients. The World Journal of Biological Psychiatry 12(8):608–619 .

Le Hebel F, Montpied P, Tiberghien A (2016) Which answering strategies do low achievers use to solve pisa science items? In: Insights from Research in Science Teaching and Learning: Selected Papers from the ESERA 2013 Conference, Springer, pp 237–252 .

Liu Z, Zhi R, Hicks A, et al (2017) Understanding problem solving behavior of 6–8 graders in a debugging game. Computer Science Education 27(1):1–29 .

MacGregor JN, Ormerod TC, Chronicle EP (1999) Spatial and contextual factors in human performance on the travelling salesperson problem. Perception 28(11):1417–1427 .

Maddox B, Bayliss AP, Fleming P, et al (2018) Observing response processes with eye tracking in international large-scale assessments: Evidence from the oecd piaac assessment. European Journal of Psychology of Education 33(3):543–558 .

Mazyavkina N, Sviridov S, Ivanov S, et al (2021) Reinforcement learning for combinatorial optimization: A survey. Computers & Operations Research 134:105400

.

McCloskey P (1985) Enhancing visuo-spatial problem solving and vjr through creative learning in functional design. Educational Review 37(3):263–280 .

McGaughey S (2007) Researchers find eye movement can affect problem-solving, cognition — beckman.illinois.edu. https://beckman.illinois.edu/news/article/2007/09/19/d0f4dbcf-5101-4bf6-b6eb-89d85702d740, [Accessed 13-10-2025] .

McInnes K, Howard J, Miles GE, et al (2009) Behavioural differences exhibited by children when practising a task under formal and playful conditions. Educational & Child Psychology 26(2) .

Meijering B, Van Rijn H, Taatgen NA, et al (2012) What eye movements can tell about theory of mind in a strategic game. PLOS ONE 8(6) .

Nasreen S, Roy AK, Guha R (2022) Exploring 'little-c'creativity through eye-parameters. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE, pp 1078–1081 .

Nasreen S, Roy AK, Guha R (2025) Exploring creativity through the eyes: A new framework using rorschach inkblot test metrics. Annals of Neurosciences p 09727531251364955 .

Newell A, Simon HA, et al (1972) Human problem solving, vol 104. Prentice-hall Englewood Cliffs, NJ .

Nisbett RE, Wilson TD (1977) Telling more than we can know: Verbal reports on mental processes. Psychological review 84(3):231 .

Nyamsuren E, Taatgen NA (2013) Set as an instance of a real-world visual-cognitive task. Cognitive Science 37(1):146–175 .

Pejić A, Molcer PS (2019) Predicting the outcome of a pisa problem solving task using strategic behavior data. In: 2019 10th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), IEEE, pp 313–318 .

Polya G (1957) How to solve it. Princeton Press .

Polya G (2014) How to solve it: A new aspect of mathematical method. In: How to solve it. Princeton university press .

Reed S (1992) Cognition: Theory and Applications. Brooks/Cole's Series in Experimental Psychology, Brooks/Cole Publishing Company, URL https://books.google.co.in/books?id=r1dEAAAAYAAJ .

Resing WC, Vogelaar B, Elliott JG (2020) Children's solving of 'tower of hanoi'tasks: dynamic testing with the help of a robot. Educational Psychology 40(9):1136–1163 .

Robertson SI (2016) Problem solving: Perspectives from cognition and neuroscience. Psychology Press .

Roy AK, Akhtar MN, Mahadevappa M, et al (2017) A novel technique to develop cognitive models for ambiguous image identification using eye tracker. IEEE Transactions on Affective Computing 11(1):63–77 .

Sailer U, Flanagan JR, Johansson RS (2005) Eye–hand coordination during learning of a novel visuomotor task. Journal of Neuroscience 25(39):8833–8842 .

Salman HA, Kalakech A, Steiti A (2024) Random forest algorithm overview. Babylonian Journal of Machine Learning 2024:69–79 .

Shaik AB, Srinivasan S (2018) A brief survey on random forest ensembles in classification model. In: International Conference on Innovative Computing and Communications: Proceedings of ICICC 2018, Volume 2, Springer, pp 253–260 .

Sinaga B, Sitorus J, Situmeang T (2023) The influence of students' problem-solving understanding and results of students' mathematics learning. In: Frontiers in Education, Frontiers Media SA, p 1088556 .

Sloman SA (1996) The empirical case for two systems of reasoning. Psychological Bulletin 119(1):3–22 .

Spivey MJ, Dale R (2011) Eye movements both reveal and influence problem solving. Oxford acedemic .

Stanovich KE, West RF (2000) Individual differences in reasoning: Implications for the rationality debate? Behavioral and Brain Sciences 23(5):645–665 .

Susac A, Bubic A, Planinic M, et al (2019) Role of diagrams in problem solving: An evaluation of eye-tracking parameters as a measure of visual attention. Physical Review Physics Education Research 15(1):013101 .

Tang H, Day E, Kendhammer L, et al (2016) Eye movement patterns in solving science ordering problems. Journal of eye movement research 9(3) .

Thomas LE, Lleras A (2007) Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. Psychonomic bulletin & review 14(4):663–668 .

Thomas LE, Lleras A (2009) Covert shifts of attention function as an implicit aid to insight. Cognition 111(2):168–174 .

Tsai MJ, Hou HT, Lai ML, et al (2012) Visual attention for solving multiple-choice science problem: An eye-tracking analysis. Computers & Education 58(1):375–385 .

Turano KA, Heidenreich SM (1999) Eye movements affect the perceived speed of visual motion. Vision Research 39(6):1177–1187 .

Udeozor C, Russo A, Glassey J, et al (2021) Exploring log data for behaviour and solution patter analysis in a serious game. Gamification and social networks in education .

Van Gog T, Paas F, Van Merriënboer JJ (2005) Uncovering expertise-related differences in troubleshooting performance: combining eye movement and concurrent verbal protocol data. Applied Cognitive Psychology 19(2):205–221 .

Wang B, Goodpaster AM, Kennedy MA (2013) Coefficient of variation, signal-to-noise ratio, and effects of normalization in validation of biomarkers from nmr-based metabonomics studies. Chemometrics and Intelligent Laboratory Systems 128:9–16 .

Wang KD, Salehi S, Arseneault M, et al (2021) Automating the assessment of problem-solving practices using log data and data mining techniques. In: Proceedings of the eighth ACM conference on learning@ scale, pp 69–76 .

Weng TS (2022) Enhancing problem-solving ability through a puzzle-type logical thinking game. Scientific Programming 2022(1):7481798 .

Wertheimer M (1945) Productive Thinking. Harper & Brothers, New York .

Wilson RC, Geana A, White JM, et al (2014) Humans use directed and random exploration to solve the explore–exploit dilemma. Journal of experimental psychology: General 143(6):2074 .

Yarkoni T, Westfall J (2017) Choosing prediction over explanation in psychology: Lessons from machine learning. Perspectives on Psychological Science 12(6):1100–1122 .

Zang S, Lin P, Chen X, et al (2022) Comparing middle school students' scientific problem-solving behavior in hands-on manipulation performance assessment: Terms by eye-tracking analysis. Scientific Programming 2022(1):6972215 .