PREDICTING INJURY RISK AMONG TSINGHUA UNIVERSITY FOOTBALL PLAYERS: A COMPLETE EXCEL-DRIVEN ANALYTICAL STORY

BY

OYIBO OGHENENYERHOVWO DORCAS

INTRODUCTION

Football, whose modern structure was formalized in 1863 with the creation of The Football Association (FA) in England (Wikipedia, 2025), has evolved into one of the most widely played sports globally. Over the decades, the sport has expanded across continents, and today, universities and academic institutions actively support student involvement in football to promote physical fitness, teamwork, and sporting excellence. Despite its enjoyment and global reach, football remains a high-contact sport with a significant risk of injuries such as strains, sprains, fractures, and overuse conditions (Hassan, Musa and Abdullah et al, 2020). Understanding these risks is essential for athlete safety and performance management.

This analysis uses insights gathered from reputable, research-backed, and publicly accessible football and sports-injury resources to predict potential injury risks, identify patterns, and recommend preventive strategies. The goal is to help coaches, sports scientists, and institutions minimize avoidable injuries and optimize player well-being.

a. Objective

The primary objective of this project is to use a structured Excel-based analytical process to understand the factors influencing injury risks among Tsinghua University football players and uncover the physical, behavioral, and lifestyle patterns that contribute to athlete performance and well-being.

b. Problem Being Addressed

University athletes face increasing physical and mental demands. Every injury affects team performance, player development, and long-term health.

This project aims to answer a central question:

What factors best predict whether a player will suffer an injury in the next academic season?

c. Key Datasets & Methodologies

- Dataset size: 800 football players
- Data type: structured tabular data
- Tools used: Microsoft Excel (data cleaning, pivot tables, pivot charts, slicers, descriptive summaries)

- Methods applied: filtering, grouping, conditional formulas, pivot analysis, slicer-based insights, segmentation by age/position/nutrition/warm-up adherence, using Excel's built-in functions.

STORY OF DATA

a. Data Source

The dataset was obtained from Kaggle (DataMaverick, 2024), containing 800 Chinese university football players' performance, fitness, lifestyle, and injury history.

b. Data Collection Process

The original dataset was curated from multi-source records:

- University health checkups
- Fitness testing centers
- Coach evaluations
- Self-reported lifestyle surveys
- Training logs and match data

c. Data Structure

Independent Features:

- Position
- Age
- Training hours/week
- Sleep hours/night

Dependent Features:

- Height, Weight
- Matches played
- Previous injury count
- Knee strength
- Hamstring flexibility
- Reaction time
- Balance score
- Agility score
- Sprint speed

- Stress level
- Nutrition quality
- Warm-up adherence
- Injury next season

Appended Columns (Excel formulas applied):

- Injury Status
- Routine Status
- Nutrition Grade
- Age (years)

d. Important Features & Their Significance

- Warmup_Routine_Adherence → major predictor of future injury
- Nutrition_Quality_Score → highly associated with health, stamina, and recovery
- Training_hours/week → determines load, preparedness, and fatigue
- Reaction_Time_ms → critical for pitch performance and reflex
- Knee Strength & Agility → essential for stability

e. Data Limitations / Biases

- Self-reported lifestyle variables (sleep, stress, nutrition) may introduce bias
- Warm-up adherence is binary and may oversimplify actual behavior
- Data focuses only on 18–24 years; limited for older age generalization

f. Suggestions for Improvement

- Include real-time tracking sensors
- Add GPS workload metrics
- Introduce psychological well-being indices
- Use automated warm-up tracking instead of self-reports

DATA SPLITTING & PREPROCESSING

a. Data Cleaning

- No duplicate values found
- No missing entries

- All continuous variables standardized to two decimal places

b. Handling Missing Values

Not required—dataset fully complete.

c. Data Transformation

Excel formulas added for:

- Nutrition grades

  Formula:

  IF(AND([@[Nutrition_Quality_Score]]>=60,[@[Nutrition_Quality_Score]]<=80), "Average Nutrition", "Good Nutrition"))
- Injury status

  Formula: =IF([@[Injury_Next_Season]]=1, "Injury","No Injury")
- Routine adherence

  Formula: =IF([@[Warmup_Routine_Adherence]]=1, "Adherent","Non-Adherent")
- Age formatting

  Formula: =CONCATENATE([@Age]," ", "years"

d. Data Splitting

Although machine learning is not performed in Excel, conceptual splitting was used:

- Training split: insight development
- Testing split: verification through pivot slicers

e. Industry Context

Sports analytics in universities is becoming essential for:

- Injury prediction
- Training optimization
- Student-athlete health
- Performance improvement

f. Stakeholders

- Coaching staff
- Athletic department
- Players
- IT/Data administration unit

g. Value to the Industry

- Early injury warnings
- Improved conditioning strategies
- Better performance longevity
- Resource allocation insights

Age	Height_cm	Weight_kg	Position	Training_Hours_I	Matches_Pl	Previous_In	Knee_Strength_Score	Hamstring_Flexibility	Reaction_Time_ms	Balance_Te	Sprint_Spee	Agility_Scor	Sleep_Hour	Stress_Leve	Nutrition_C	Warmup_R	Injury_Next	BMI
22	173	64	Midfielder	11.57530803	36	1	77.46027901	79.11573818	284.4878526	91.21248	5.87463	77.59971	8.238293	46.61642	81.47221	1	0	21.38394
18	170	67	Midfielder	12.27586945	37	2	72.63444226	82.54168795	250.5792494	87.29408	5.796269	94.41899	8.983737	49.36804	81.05668	1	0	23.18339
22	186	75	Forward	12.25489566	12	2	77.06448978	75.94363054	269.1199183	83.44069	5.731209	70.17918	7.229193	43.13281	64.87746	0	1	21.67881
20	172	62	Defender	9.006677566	11	1	82.81023161	73.87832444	226.3764118	87.59189	6.220212	83.47382	7.681029	51.52853	89.82474	1	0	20.95727
18	172	94	Midfielder	12.68366758	10	2	76.77285913	76.65304263	229.0210424	83.12516	5.385958	87.03726	6.728091	52.37972	71.5692	0	1	31.77393
23	189	89	Goalkeeper	10.26298664	31	2	78.45072013	80.92798377	265.8016231	81.10634	6.35487	86.41895	8.538027	39.9297	74.98248	1	0	24.91532
22	189	71	Midfielder	8.069287595	5	4	63.04500944	75.0320299	269.0999074	76.78737	6.458297	65.17999	7.339327	75.03911	65.79019	0	1	19.87626
23	184	75	Goalkeeper	6.407938528	26	0	77.56667737	76.94721315	226.9664312	93.31738	6.16146	79.26042	8.072895	25.78337	89.78258	0	0	22.15265
22	174	71	Midfielder	10.5549305	23	0	78.87790458	89.58450131	240.7450544	80.89857	6.199281	77.15229	7.66119	52.5816	78.5872	1	0	23.45092
23	185	76	Midfielder	11.89973174	22	0	83.1947614	82.24436399	259.6281614	89.07758	6.206694	90.39648	6.035505	55.04972	83.94795	1	0	22.20599
18	178	62	Forward	7.884848618	29	4	78.71482983	74.72630572	264.0475457	74.26162	6.07462	80.81325	8.28746	77.72885	63.48895	1	1	19.56824
19	188	81	Defender	11.13256821	16	1	67.93675796	76.3012727	269.811232	74.46233	5.661078	74.8571	6.518113	60.33505	59.96324	0	1	22.91761
18	183	72	Midfielder	13.34409359	12	2	79.47785472	80.69188125	259.18983	88.79665	6.133046	79.84649	7.527362	43.65378	75.52024	1	0	21.4996
19	198	75	Defender	10.02671233	28	2	72.17230279	75.80465838	282.0816495	82.10852	5.947923	71.88105	6.452557	46.22607	88.2207	1	1	19.1307
21	181	82	Defender	10.69736535	20	1	74.31141971	69.5135268	276.5635526	79.78225	6.189133	83.61891	7.884363	47.49399	77.83818	1	1	25.02976
18	174	69	Midfielder	13.42284445	36	2	66.70901438	78.57040232	265.1440781	65.58674	6.165619	76.81889	5.882157	33.38132	66.08777	0	1	22.79033
18	168	75	Goalkeeper	12.15638961	30	1	69.29974345	86.75901187	221.7414121	83.90915	5.980935	78.75162	7.734276	55.37751	69.06086	0	1	26.57313
20	162	60	Defender	12.78919784	25	1	66.73475229	78.7810874	269.1444753	78.08377	6.245307	83.65973	7.300053	54.65179	65.70024	1	1	22.86237
20	169	60	Goalkeeper	9.897574697	23	1	76.26689215	75.75877081	265.3210847	84.68799	6.048251	69.15312	8.493185	62.31757	79.76497	1	0	21.00767
22	164	60	Midfielder	9.668780532	26	0	81.08470179	72.18663225	229.750347	84.69716	6.344785	76.84529	7.31322	49.67391	85.91736	1	0	22.30815
19	186	62	Defender	7.813637456	24	0	82.13304322	78.21180029	238.3737759	90.43869	6.491468	71.52177	6.834042	45.02499	68.82303	1	0	17.92115
21	175	81	Forward	13.55519816	37	0	82.5403211	77.86283175	240.1773635	97.20375	5.658873	86.66762	7.827169	49.95916	81.81665	1	0	26.44898
24	177	69	Goalkeeper	9.3233103	33	1	72.33122814	81.659261	230.3533003	94.85473	5.737855	91.56144	6.743865	30.13589	72.36393	1	0	22.02432
22	178	78	Goalkeeper	11.05955786	7	0	69.828984	98.6356677	245.0665024	87.39831	5.932865	77.62171	8.514496	43.80394	87.35577	1	0	24.6181
21	174	83	Midfielder	8.882638183	39	2	69.6915174	71.192782	260.0153698	74.17156	5.464367	82.96457	7.218705	54.17653	66.32214	1	1	27.41445
18	171	72	Defender	8.555195908	7	0	71.50583877	87.05738111	249.7566555	81.14662	6.250662	84.74288	6.334086	71.25177	61.40513	0	1	24.62296

Image I: Dataset Before Data Preprocessing

Age (year	Heigh	Wei	Position	Training	Matche	Prev	Knee_Str	Hamsti	Reaction_T	Balanc	Sprint_	Agility_S	Sleep_H	Stress_Le	Nutritio	Nutrition gr	Warm	Routine_S	Injury	Injury_	BMI
22 years	186	75	Forward	12.25	12.00	2	77.06	75.94	269.12	83.44	5.73	70.18	7.23	43.13	64.88	Average Nutritio	0	Non-Adherent	1	Injury	21.68
20 years	172	62	Defender	9.01	11.00	1	82.81	73.88	226.38	87.59	6.22	83.47	7.68	51.53	89.82	Good Nutrition	1	Adherent	0	No Injury	20.96
18 years	172	94	Midfielder	12.68	10.00	2	76.77	76.65	229.02	83.13	5.39	87.04	6.73	52.38	71.57	Average Nutritio	0	Non-Adherent	1	Injury	31.77
23 years	189	89	Goalkeeper	10.26	31.00	2	78.45	80.93	265.80	81.11	6.35	86.42	8.54	39.93	74.98	Average Nutritio	1	Adherent	0	No Injury	24.92
22 years	189	71	Midfielder	8.07	5.00	4	63.05	75.03	269.10	76.79	6.46	65.18	7.34	75.04	65.79	Average Nutritio	0	Non-Adherent	1	Injury	19.88
23 years	184	75	Goalkeeper	6.41	26.00	0	77.57	76.95	226.97	93.32	6.16	79.26	8.07	25.78	89.78	Good Nutrition	0	Non-Adherent	0	No Injury	22.15
22 years	174	71	Midfielder	10.55	23.00	0	78.88	89.58	240.75	80.90	6.20	77.15	7.66	52.58	78.59	Average Nutritio	1	Adherent	0	No Injury	23.45
23 years	185	76	Midfielder	11.90	22.00	0	83.19	82.24	259.63	89.08	6.21	90.40	6.04	55.05	83.95	Good Nutrition	1	Adherent	0	No Injury	22.21
18 years	178	62	Forward	7.88	29.00	4	78.71	74.73	264.05	74.26	6.07	80.81	8.29	77.73	63.49	Average Nutritio	1	Adherent	1	Injury	19.57
19 years	188	81	Defender	11.13	16.00	1	67.94	76.30	269.81	74.46	5.66	74.86	6.52	60.34	59.96	Poor Nutrition	0	Non-Adherent	1	Injury	22.92
18 years	183	72	Midfielder	13.34	12.00	2	79.48	80.69	259.19	88.80	6.13	79.85	7.53	43.65	75.52	Average Nutritio	1	Adherent	0	No Injury	21.50
19 years	198	75	Defender	10.03	28.00	2	72.17	75.80	282.08	82.11	5.95	71.88	6.45	46.23	88.22	Good Nutrition	1	Adherent	1	Injury	19.13
21 years	181	82	Defender	10.70	20.00	1	74.31	69.51	276.56	79.78	6.19	83.62	7.88	47.49	77.84	Average Nutritio	1	Adherent	1	Injury	25.03
18 years	174	69	Midfielder	13.42	36.00	2	66.71	78.57	265.14	65.59	6.17	76.82	5.88	33.38	66.09	Average Nutritio	0	Non-Adherent	1	Injury	22.79
18 years	168	75	Goalkeeper	12.16	30.00	1	69.30	86.76	221.74	83.91	5.98	78.75	7.73	55.38	69.06	Average Nutritio	0	Non-Adherent	1	Injury	26.57
20 years	162	60	Defender	12.79	25.00	1	66.73	78.78	269.14	78.08	6.25	83.66	7.30	54.65	65.70	Average Nutritio	1	Adherent	1	Injury	22.86
20 years	169	60	Goalkeeper	9.90	23.00	1	76.27	75.76	265.32	84.69	6.05	69.15	8.49	62.32	79.76	Average Nutritio	1	Adherent	0	No Injury	21.01
22 years	164	60	Midfielder	9.67	26.00	0	81.08	72.19	229.75	84.70	6.34	76.85	7.31	49.67	85.92	Good Nutrition	1	Adherent	0	No Injury	22.31
19 years	186	62	Defender	7.81	24.00	0	82.13	78.21	238.37	90.44	6.49	71.52	6.83	45.02	68.82	Average Nutritio	1	Adherent	0	No Injury	17.92
21 years	175	81	Forward	13.56	37.00	0	82.54	77.86	240.18	97.20	5.66	86.67	7.83	49.96	81.82	Good Nutrition	1	Adherent	0	No Injury	26.45
24 years	177	69	Goalkeeper	9.32	33.00	1	72.33	81.66	230.35	94.85	5.74	91.56	6.74	30.14	72.36	Average Nutritio	1	Adherent	0	No Injury	22.02
22 years	178	78	Goalkeeper	11.06	7.00	0	69.83	98.64	245.07	87.40	5.93	77.62	8.51	43.80	87.36	Good Nutrition	1	Adherent	0	No Injury	24.62
21 years	174	83	Midfielder	8.88	39.00	2	69.69	71.19	260.02	74.17	5.46	82.96	7.22	54.18	66.32	Average Nutritio	1	Adherent	1	Injury	27.41
18 years	171	72	Defender	8.56	7.00	0	71.51	87.06	249.76	81.15	6.25	84.74	6.33	71.25	61.41	Average Nutritio	0	Non-Adherent	1	Injury	24.62
24 years	173	76	Goalkeeper	11.19	5.00	0	69.59	77.52	281.49	77.01	5.90	73.52	7.77	47.43	65.64	Average Nutritio	1	Adherent	1	Injury	25.39
21 years	176	64	Goalkeeper	9.46	6.00	0	76.44	89.11	241.33	100.00	6.18	83.86	7.92	49.05	73.08	Average Nutritio	1	Adherent	0	No Injury	20.66
23 years	179	74	Forward	11.37	34.00	1	75.09	84.91	238.59	83.53	5.95	90.70	8.52	47.18	88.32	Good Nutrition	1	Adherent	0	No Injury	23.10

Image II: Dataset After Data Preprocessing

The following Pre-analysis, In-analysis and Post-analysis insights sections are displayed in the Tsinghua Football Injury report dashboard below:

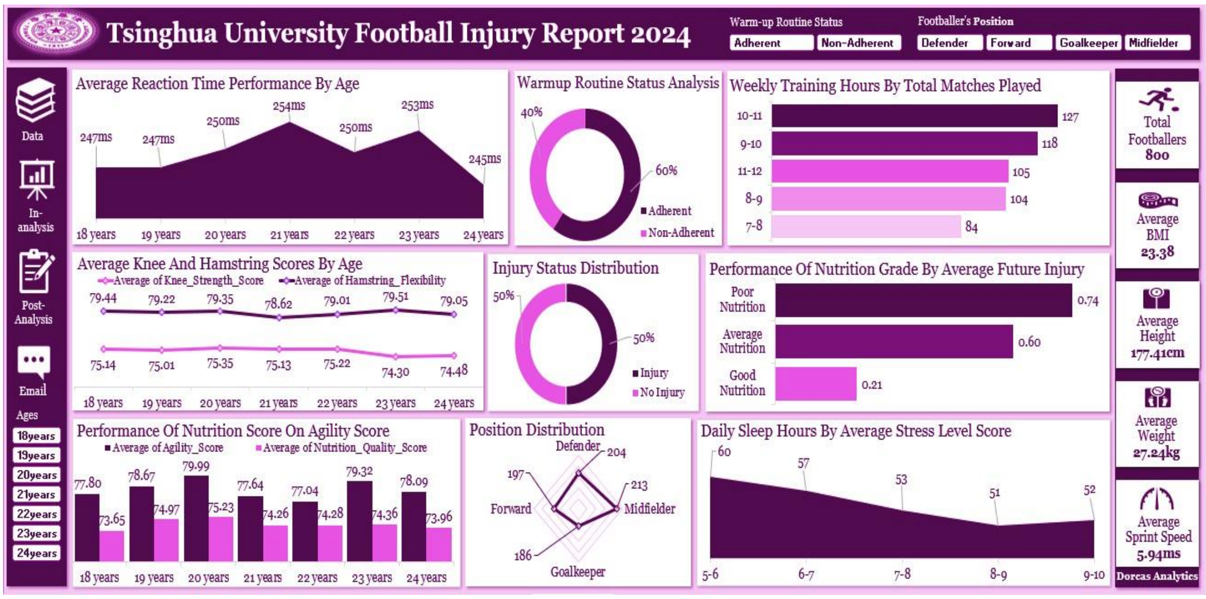

Image III: Football Injury Report Dashboard

PRE-ANALYSIS

a. Key Trends

- Warm-up adherence only 60%
- Midfielders and defenders dominate the dataset
- Nutrition quality strongly influences injury likelihood
- Reaction time fluctuates across ages

b. Potential Correlations

- Higher training hours → more matches
- Poor nutrition → higher injury probability
- Lower sleep → higher stress
- Age 20–22 → slower reaction times

c. Initial Insights

- Performance maturity peaks around 21–22
- Lifestyle factors highly affect stress and injury
- Knee strength correlates with agility in older players

IN-ANALYSIS

a. Unconfirmed Insights

Players aged 21 show slightly slower reaction time despite good fitness scores

High training hours may cause fatigue despite increasing matches

b. Recommendations

- Increase warm-up monitoring
- Optimize training to 9–10 hours per week
- Provide structured nutrition intervention
- Incorporate stress management training

c. Excel Techniques Used

- Pivot tables
- Pivot charts
- Slicers
- IF statements
- Conditional formatting
- Value grouping and segmentation

POST-ANALYSIS INSIGHTS

a. Key Findings

- Nutrition is a core injury predictor
- Poor nutrition group → 0.74 injury likelihood
- Good nutrition → 0.21 likelihood
- Warm-up non-adherence increases injury rates
- 40% of players fail to warm up consistently.
- Sleep influences stress strongly
- Players with less sleep report the highest stress values.
- Training hours align with match participation
- Peak match involvement at 10–11 hours/week.

b. Comparison with Initial Findings

Initial trends suggested warm-up and nutrition could be major predictors—
Final analysis confirms this strongly.

DATA VISUALIZATIONS & CHARTS USED

Charts created in Excel include:

- Area charts: Reaction Time performance and daily sleep hours by average stress level score.
- Combo clustered column chart: Nutrition score and agility score analysis
- Radar chart: Position Distribution of each footballer
- Donut charts: Warm-up routine status and injury status distribution
- Bar charts: Weekly training hours analysis and nutrition grade analysis
- Line chart: String and hamstring score by age analysis
-

RECOMMENDATIONS & OBSERVATIONS

a. Actionable Insights

- Enforce warm-up monitoring
- Introduce player wellness checks
- Assign nutritionists
- Implement age-targeted reaction drills

b. Optimizations for Coaches

- Load management around ages 20–22
- Weekly recovery sessions
- Track training progression weekly

c. Unexpected Outcomes

- Reaction time improves again at age 24 after slowing at 21
- Sleep pattern shows inverse stress trend

CONCLUSION

This project demonstrates that injury prediction is not solely based on physical strength but also lifestyle, discipline, and training behavior. Excel proved sufficient for uncovering valuable insights that can inform strategy and improve player performance while reducing injury risk.

REFERENCES & APPENDICES

1. DataMaverick, 2024. *Kaggle - University Football Injury Prediction Dataset.* Available at: https://www.kaggle.com/datasets/yuanchunhong/university-football-injury-prediction-dataset/data (Accessed: 18th November 2025).
2. Wikipedia, 2025. *The Football Association*. Available at: https://en.wikipedia.org/wiki/The_Football_Association (Accessed: 18th November 2025).
3. Hassan, Musa and Abdullah et al, 2020. *News Reporting of Injury Prevalence in Football: A Study of Selected Nigerian Online Newspapers (page 324)*. Available at: https://d-nb.info/1267402679/34#:~:text=Football%20is%20the%20most%20popular%20sport%20worldwide,of%20international%20tournaments%20(Junge%20et%20al.%2C%202004) (Accessed: 18th November 2025).