MATHEMATICAL AND STATISTICAL ANALYSIS OF HIGHER EDUCATION IN MIZORAM

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DAVID ROSANGLIANA

MZU REGISTRATION NUMBER : 62 of 2011
PH.D. REGISTRATION NUMBER : MZU/Ph.D./742 of 11.05.2015

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
SCHOOL OF PHYSICAL SCIENCES
JUNE, 2023

MATHEMATICAL AND STATISTICAL ANALYSIS OF HIGHER EDUCATION IN MIZORAM

BY

David Rosangliana
Department of Mathematics and Computer Science

Supervisor: Prof. Jamal Hussain

Submitted

In partial fulfillment of the requirement of the Degree of Doctor of Philosophy in Mathematics of Mizoram University, Aizawl

Mizoram University
TANHRIL : AIZAWL – 796004

THESIS CERTIFICATE

This is to certify that the research thesis entitled ***Mathematical and Statistical Analysis of Higher Education in Mizoram*** submitted by ***David Rosangliana*** to Mizoram University, Tanhril, Aizawl, for the award of the degree of Doctor of Philosophy is a bonafide record of research work carried out by him under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

(PROF. JAMAL HUSSAIN)
Research Supervisor

Place : Aizawl
Date: 06.06.2023

DECLARATION OF THE CANDIDATE

I, David Rosangliana, hereby declare that the subject matter of this thesis entitled **"Mathematical and Statistical Analysis of Higher Education in Mizoram"** is the record of the work done by me, that the contents of this thesis did not form basis of the award of any previous degree to me or to the best of my knowledge to anybody else, and that the thesis has not been submitted by me for any research degree in any other University/Institute.

This is being submitted to the Mizoram University for the degree of Doctor of Philosophy in Mathematics.

Dated: 06.06.2023

(DAVID ROSANGLIANA)
Candidate

(**PROF. M. SUNDARARAJAN**)
Head
Department of Mathematics & Computer Science

(PROF. JAMAL HUSSAIN)
Supervisor

ACKNOWLEDGEMENTS

I express my sincere thanks and gratitude to my guide and supervisor Jamal Hussain, Professor, Department of Mathematics and Computer Science, Mizoram University who has been a source of inspiration and a guide on every step during my entire Ph.D. course. I also want to thank my friend Vanlalruata, a research scholar at the Mizoram University's department of mathematics and computer science, without whose help it would be very difficult to do this assignment.

I am also grateful to all the faculty members and research scholars of the Department of Mathematics & Computer Science, Mizoram University for their support, academic discussions and insightful input.

I thank Govt. of Mizoram authority for giving me permission to pursue my research, where permission is needed for college teacher. I express my heartfelt gratitude to University Grants Commission – NERO for providing me assistance through Teacher Fellowship under Faculty Development Programme (Vide Order no.F.7-2/TF/2012/NERO/167 dated 20th January 2017)

My fellow research students at Mizoram University's Mathematics and Computer Science department have all been incredibly kind and helpful to me during my study; they have all been there for me when I've needed them most.

I would also like to thank Dr. C. Zorammuana, Head, Department of Mathematics, Lunglei Govt. College, Dr. Samuel Lalmuanawma, Assistant Professor, Department of Computer Science, Govt. Champhai College and Dr. Lalhmangaihzuala, Assistant Professor, Department of Mathematics, Govt. Serchhip College for their kind suggestions and encouragement. Help from Dr. Samuel Lallianrawna, Assistant Professor, Department of Chemistry, Govt. Zirtiri Residential Science College, Dr. Lalmalsawma Khiangte, Associate Professor, Department of Political Science, Govt. Aizawl College and Lalnunmawii, Assistant Professor, Department of Commerce, Govt. Aizawl College was also a great source of inspiration for me.

Support and help I received from Dr. Benjamin L. Saitluanga, Professor, Department of Geography and Resource Management, Mizoram University is worthwhile for my research.

I express my deep sense of gratefulness to my father Revd. H. Lalchhanhima and my mother PC Lalramliani who had always supported, encouraged and helped me throughout this course of study so that I could give more attention to my studies.

My life is filled with joy thanks to my wife, Lalbiakzami (Nunui), and my four children, Biaksangpuia Hnamte, Laldinsangi Hnamte, Zothanzuala Hnamte and Lalrinzuali Hnamte. They are the inspiration for all of my accomplishments and the source of my happiness.

Above all, I would like to express my heartiest thanks to the Almighty God for blessing me with health, strength and guidance in order to overcome many difficulties and problems during my study and enabling me to complete my work.

Date: 06.06.2023 (DAVID ROSANGLIANA)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AISHE	All India Survey on Higher Education
ACF	Auto Correlation Function
ADF	Augmented Dickey-Fuller
AI	Artificial Intelligence
AICTE	All India Council for Technical Education
ANN	Artificial Neural Network
AR	Auto Regressive
ARIMA	Auto Regressive Integrated Moving Average
DHE	Department of Higher Education
DL	Deep Learning
GER	Gross Enrolment Ratio
GPI	Gender Parity Index
HS	High School
KNN	K Nearest Neighbour
LOOCV	Leave-One-Out Cross-Validation
LSTM	Long Short Term Memory
MA	Moving Average
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MHRD	Ministry of Human Resource Development
MHRD	Ministry of Human Resource Development
ML	Machine Learning
MOE	Ministry of Education
MS	Middle School
MSE	Mean Square Error
NAAC	National Assessment and Accreditation Council

NBA	National Board of Accreditation
NEP	National Education Policy
NSSO	National Sample Survey Organization
PACF	Partial-Autocorrelation Function
PG	Post Graduate
Ph.D	Doctor of Philosophy
PS	Primary School
PTR	Pupil Teacher Ratio
RMSE	Root Mean Square Error
RTE	Right to Education
SD	Standard Deviation
TSA	Time Series Analysis
UDISE	Unified District Information System for Education
UG	Under Graduate
UGC	University Grants Commission
UNESCO	United Nations Educational, Scientific and Cultural Organization
UPS	Upper Primary School
VAR	Vector Auto Regression
WOA	Whale Optimization Algorithm
WOASVR	Whale Optimization Algorithm and Support Vector Regression

CHAPTER 1

INTRODUCTION

1.0 Introduction

Education is one of the fundamental factors in the overall development of a nation. It is a constant source of lifelong learning and has a direct bearing on human capital, which is the nation’s most valuable and crucial resource. The wealth and growth of any nation can only be achieved through the use of human capital. Humans are the active agents who can raise capital, construct political, economic, and social organizations, and properly utilize natural resources, leading to national development. Capital and natural resources are passive factors of production (Türkkahraman, 2012). Education influences not only the educated but also the community. In other words, education and educational institutions at all levels must raise sufficient numbers of competent people to provide a more successful society. Education is vital to a nation's development, as it provides crucial knowledge and skills for economic growth and overall progress. Education is a critical element for economic and human resource development (Kumar, 2012).

Higher education is extremely important for the country since it is a potent weapon in the development of a knowledge-based society in the twenty-first century. High-quality higher education is no longer considered a luxury but rather a necessity for national, social, and economic growth, according to the United Nations Educational, Scientific, and Cultural Organization (UNESCO). The overarching goal of education is to produce a significant number of educated men and women who are capable of comprehending the world and effecting change that will result in adequate health and education services, a better environment, and the elimination of ignorance and deprivation (limitations) that continue to suffocate developing societies today. (Rao *et.*

al., 2004) argue for a policy that emphasises the education of those who are underprivileged and live in poverty, following the principles of equality, quality, and efficiency. To keep up with the rapid expansion of higher education in both developed and developing countries, there is a pressing need to improve the quantity and quality of higher education institutes in the country. Higher education is critical to a country's progress. It is vital for economic and social growth as well as shaping human resources, as it allows people to respond to significant socioeconomic, cultural, moral, and spiritual challenges. It has traditionally been acknowledged as a major tool for achieving national social, economic, and political goals. Higher education is the main tool for assuring people's and countries' advancement. Educated persons can think critically, assess problems in society, find answers to those problems, apply relevant solutions, and assume societal duties. Higher education supplies ideas and men to influence the future and to support all other levels of education, including kindergarten. It is, after all, the backbone of an economy and a potent tool for achieving national goals (Bhutia, 2005).

Indian higher education has grown dramatically since Independence. Compared to 1947, when there were just 21 universities and 420 affiliated colleges (Yadav, 1985), there were 1043 universities, 42343 colleges, and 11779 stand-alone institutions in 2019. (AISHE, 2019-20). There are 522 general, 177 technical, 63 agriculture and allied, 66 medical, 23 law, 12 Sanskrit, and 11 language universities in 2019, with the remaining 145 universities falling into other categories. Consequently, the enrolment of students has increased from a mere 1 lakh in 1950 (Thorat, 2010) to over 38.5 million, with 19.6 million boys and 18.9 million girls. About 79.5% of the students are enrolled in undergraduate programs and 2,02,550 students are enrolled in Ph.D. programs, which is about 0.5% of the total student enrolment. Undergraduate students are most likely to be enrolled in arts, humanities, and social sciences (32.7%), followed by science (16%), commerce (14.9%), and engineering and technology (12.6%). 16.6% of the colleges have enrolments less than 100, and only 4% have enrolments greater than 3000 (AISHE, 2019–20). Although India has one of the largest higher education systems in the world, the gross enrolment ratio (GER) in higher education is still lower than the average of some of the developing nations and developed countries.

In fact, GER is the number of students enrolled in a given level of education regardless of age, divided by the population of the official age group that corresponds to the given level of education, and multiplied by 100 (http://uis.unesco.org/en/glossary-term//gross-enrolment-ratio). Looking at the trends of growth in higher education in India, there has been an increase in terms of different types of institutions and enrolment. If, on the other hand, we looked at the GER of various countries, we would have to concede that India is still much behind the curve, particularly when compared to developed countries. In India, the GER in higher education is 27.1, calculated for the 18-23 age groups. The male population has a GER of 26.9, while the female population has a GER of 27.3. For scheduled castes, it is 23.4, and for scheduled tribes, it is 18.0, as compared to the national GER of 27.1 (AISHE, 2019-20).

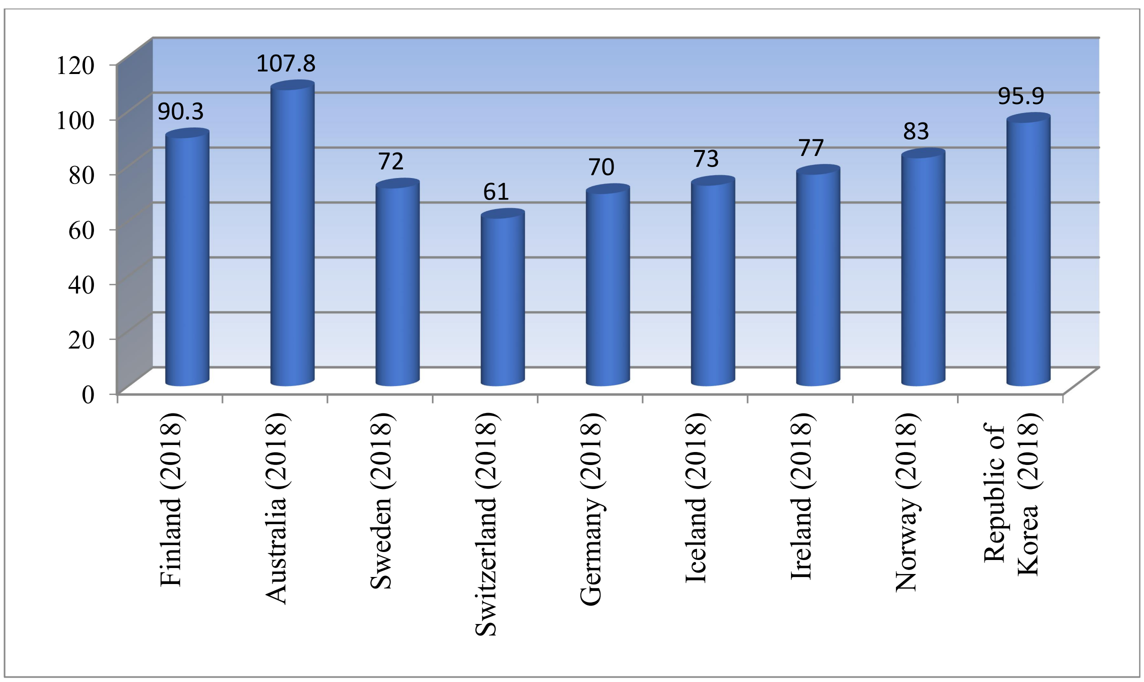

Figure 1.1: Higher education GER of some developed countries

Source: data.uis.unesco.org (data as on 30th September 2020)

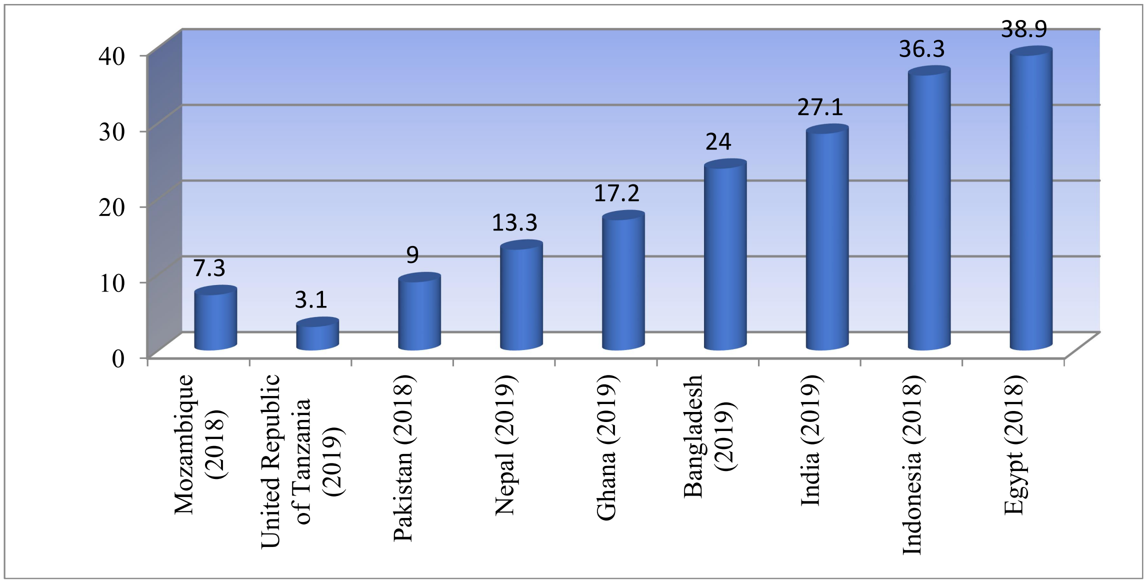

Figure 1.2: Higher education GER of some developing countries

Source: data.uis.unesco.org (data as on 30th September 2020)

With the growing size and diversity of the higher education sector, particularly in terms of courses, management, and geographical coverage, it has become necessary to develop a sound database on higher education. This is also required for planning, policy formulation, fulfilling international commitments, research, etc. To address this issue, the Department of Higher Education (DHE) and the Ministry of Human Resource Development (MHRD) decided to conduct the All India Survey on Higher Education (AISHE) on an annual basis beginning in 2010-11 with the following goals: to identify and capture all institutions of higher learning in the country and to collect data on various aspects of higher education from all higher education institutions. In India, the government has set a target of increasing the GER from its present level of about 21.1% (AISHE 2012-13 Report) to 35% by the year 2035. A reliable and comprehensive database is an immediate requirement to measure the actual GER, and efforts need to be taken to improve it. For the purpose of this survey, "higher education" has been defined as the education that is obtained after completing 12 years of schooling or its equivalent and is of the duration of at least nine months (full time) or after completing 10 years of schooling and is of the duration of at least three years.

Access to higher education can open better employment and income opportunities to all people, especially the underprivileged sections of society. Among various levels of education, higher education has a persistent and powerful impact on the development of a nation. It empowers the individual with the necessary skills and competence for achieving personal and social goals, which in turn contribute to the development of the nation (Kale, 2006). Since 1950, access to higher education has been routine for all sections of society. However, only since the IX Plan has the Government of India made access, equity, and excellence the major pillars of the development of higher education in the country. It has resulted in good growth in enrolment since then (Thangaraj, 2016). Over the last two decades, India has remarkably transformed its higher education landscape. It has created widespread access to low-cost, high-quality university education for students of all levels. With well-planned expansion, India has not only bettered its enrolment but has dramatically enhanced its learning outcomes (FICCI Higher Education Summit, 2013). In 1947, the GER in India, worked out on the basis of the relevant age group of 18–23 years, was just 0.7 (Singh and Ahmad, 2011), but in 2019–20 it has increased to 27.1. The GER for the male population is 26.9, and for the female population, it is 27.3. In 2019, 18 of the 37 states and union territories had a higher GER than the national GER of 27.1%. Fortunately, in 2019, among the Indian States and UTs, 15 States and UTs reached more than 30% of GER. It can be assumed that India is now ready to compete in a globalized world and in the emerging knowledge society in terms of its access to higher education (AISHE, 2019–20). As recommended by National Education Policy 2020, higher education institutions may be free from rigid boundaries between sciences, commerce, engineering, the arts, and the humanities. There should be flexibility, variety, and balance between courses that will meet the demands of society; no particular course needs to be overemphasized. By increasing access to higher education, India will be able to play a major role in leveraging its fast-growing global demographic and emerge as a knowledge-creating country as well as a knowledge provider. By setting up new institutions in a planned manner, India may be able to bridge the regional imbalances and disparities across different disciplines and be in a position to address the economic, social, and technological needs of the country. Further, traditional education should be supplemented with skill-based studies. This would significantly raise the GER, bringing it in line with the global average (Jeelani, 2012). Besides, to increase access to higher education; there must be cooperation between the central and state governments and joint efforts with the private sector. Higher education is no longer a luxury, but has become one of the principal means of attaining upward social mobility. The access rate to higher education is still quite low, and with the ever-increasing population, the gap between the number of college and university seats available and the requirements seems to be widening. To cater to the growing demand for higher education and ensure accessibility for the growth of young people, a massive expansion plan needs to be undertaken (Vyas & Basu, 2009).

The status of higher education in Indian states and union territories in terms of number of institutions, GER and PTR is depicted in the following Table 1.1.

TABLE 1.1: Institution numbers, GER & PTR of Indian States and Union Territories

Sl. No.	States / Union Territories	No. of University	No. of College	No. of Stand-alone Inst.	GER	PTR
1	Andaman & Nicobar Islands	0	8	1	20.0	22
2	Andhra Pradesh	41	2750	843	35.2	19
3	Arunachal Pradesh	10	39	13	35.4	28
4	Assam	26	558	92	17.3	28
5	Bihar	35	874	166	14.5	59
6	Chandigarh	3	25	10	52.1	28
7	Chhatisgarh	28	810	75	18.5	26
8	Dadra & Nagar Haveli	0	8	2	9.4	27
9	Daman & Diu	0	10	2	6.1	14

10	Delhi	28	179	112	48.0	52
11	Goa	3	58	12	28.4	15
12	Gujarat	76	2275	308	21.3	26
13	Haryana	53	1087	252	29.3	24
14	Himachal Pradesh	27	344	83	40.8	27
15	Jammu & Kashmir	15	316	118	32.4	35
16	Jharkhand	32	323	91	20.9	60
17	Karnataka	69	4047	1690	32.7	15
18	Kerala	23	1417	452	38.8	18
19	Ladakh	1	5	4	7.9	15
20	Lakshadweep	0	0	-	7.5	12
21	Madhya Pradesh	66	2411	379	24.2	34
22	Maharashtra	65	4494	2393	32.3	26
23	Manipur	8	102	29	38.9	23
24	Meghalaya	10	67	21	26.1	24
25	**Mizoram**	**3**	**35**	**16**	**26.1**	**17**
26	Nagaland	5	67	20	18.5	18
27	Odisha	32	1087	386	21.7	25
28	Puducherry	4	79	14	46.3	13
29	Punjab	32	1079	421	28.2	17
30	Rajasthan	89	3380	567	24.1	29
31	Sikkim	8	22	8	75.8	34
32	Tamil Nadu	59	2610	914	51.4	17
33	Telangana	24	2071	514	35.6	17
34	Tripura	4	53	12	20.2	36
35	Uttar Pradesh	81	7788	1116	25.3	40
36	Uttarakhand	36	454	166	41.5	27
37	West Bengal	47	1411	477	19.9	33
	All India	**1043**	**42343**	**11779**	**27.1**	**26**

Source: AISHE Report, 2019-20

In primary, secondary, and higher education, both quality and quantity are vital, and they must be maintained in tandem to be effective. A variety of steps have been implemented to improve the quality of education by the Indian government, state governments, the University Grants Commission (UGC), the All India Council for Technical Education (AICTE), and other organizations. The UGC invests a large sum of money to run UGC Human Resource Development Centers throughout India. The National Assessment and Accreditation Council (NAAC) and the National Board of Accreditation (NBA) are two organizations that play a key role in evaluating higher education institutions. The current status of institutions accredited by NAAC is given below:

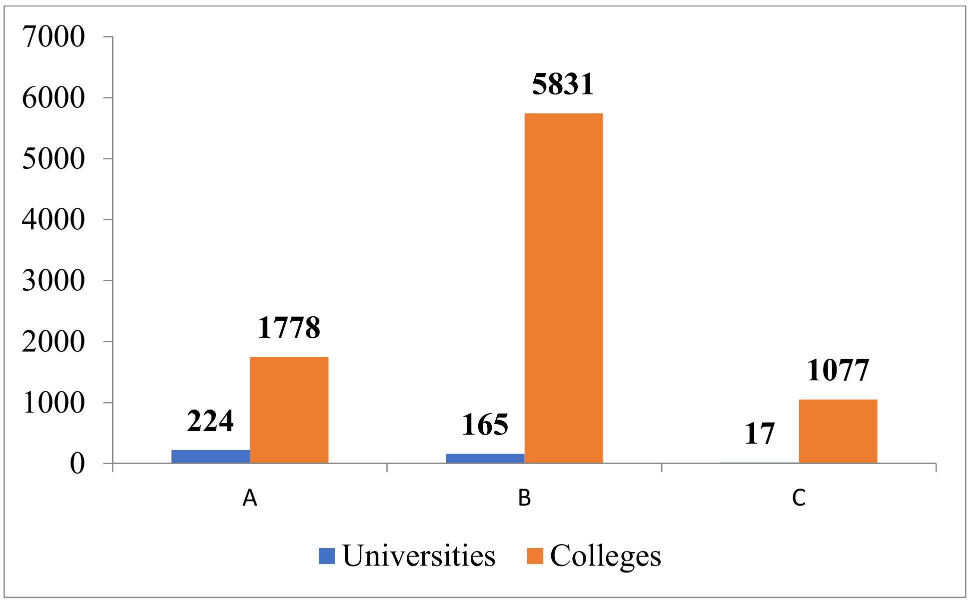

Figure 1.3: Grade Break up of Institutions accredited by NAAC (as on 21/06/2022)

Source: naac.gov.in

Out of 406 Universities having valid accreditation from NAAC, 224 have secured A grade and there are 165 and 17 Universities securing B and C grade respectively. Out of 8686 accredited colleges, 1778, 5831 and 1077 colleges secure A, B and C grade respectively.

Table 1.2: State wise number of Universities and Colleges accredited by NAAC

Sl. No.	State	Universities	Colleges	Total
1	Andaman and Nicobar Islands	0	3	**3**
2	Andhra Pradesh	15	374	**389**
3	Arunachal Pradesh	3	8	**11**
4	Assam	8	213	**221**
5	Bihar	8	164	**172**
6	Chandigarh	2	15	**17**
7	Chhattisgarh	8	177	**185**
8	Dadra and Nagar Haveli	0	3	**3**
9	Daman & Diu	0	2	**2**
10	Delhi	18	88	**106**
11	Goa	1	25	**26**
12	Gujarat	23	492	**515**
13	Haryana	16	335	**351**
14	Himachal Pradesh	9	71	**80**
15	Jammu and Kashmir	7	103	**110**
16	Jharkhand	7	121	**128**
17	Karnataka	33	881	**914**
18	Kerala	8	285	**293**
19	Ladakh	0	3	**3**
20	Madhya Pradesh	17	326	**343**
21	Maharashtra	35	1834	**1869**
22	Manipur	1	33	**34**
23	Meghalaya	2	22	**24**
24	Mizoram	1	25	**26**
25	Nagaland	1	34	**35**
26	Odisha	14	290	**304**
27	Puducherry	2	25	**27**
28	Punjab	10	281	**291**
29	Rajasthan	31	265	**296**
30	Sikkim	2	8	**10**
31	Tamil Nadu	43	829	**872**
32	Telangana	15	259	**274**
33	Tripura	2	22	**24**
34	Uttar Pradesh	37	604	**641**
35	Uttarakhand	11	65	**76**
36	West Bengal	16	401	**417**
	All India	**406**	**8686**	**9092**

Source: naac.gov.in; as on 21/06/2022

1.1 The study area: Mizoram

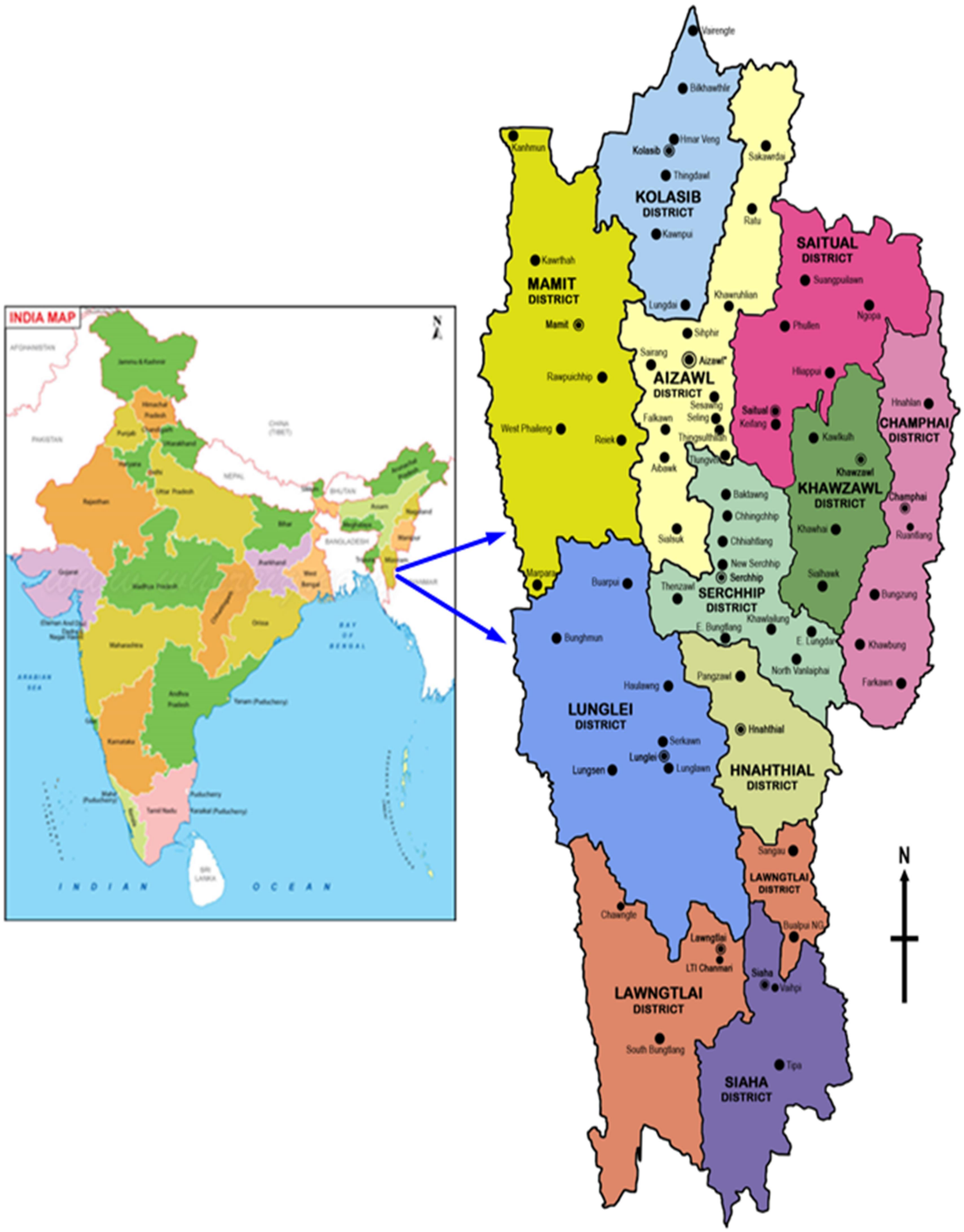

Figure 1.4: Map of Mizoram state

Mizoram is one of the northeast states in India, sandwiched between Bangladesh in the west and Myanmar in the east. It is bounded in the north by the state of Assam and Manipur. Mizoram was an excluded area during the British colonial period. After Indian independence in 1947, it was one of the districts of Assam state till it attained the status of a Union Territory (UT) on 21st January, 1972. According to the 1971 census, the population of Mizoram was 3,32,390 with a literacy percentage of 53.79%. It attained statehood on 20th February, 1987. It shares 404 kilometres long international borders with Myanmar and 318 kilometres with Bangladesh. The geographical location of Mizoram is $92^{0}.15'$ E to $93^{0}.29'$ E Longitude and $21^{0}.58'$ N to $24^{0}.35'$ N Latitude. Aizawl is the capital city of Mizoram. There are 11 Districts, 23 Sub-Divisions and 26 RD Blocks. Its area covers 21,081 square kilometres. According to 2011 census, there are 830 villages, 2,22,853 households, 10,97,206 populations and a literacy rate of 91.33% (Mizoram Statistical Handbook, 2020).

Mizoram is growing and making strides forward in higher education in comparison to other states in the region. On the 15th of August 1958, the state's first higher education institution, Aijal College, which is now known as Pachhunga University College, was started. As the Indian government works to expand higher education opportunities throughout the country, states like Mizoram have a long way to go and must move quickly to catch up with the rest of the country.

At present, Mizoram has only three universities or university-level institutions, thirty-five colleges, and sixteen stand-alone institutions. It is important to study and analyse the status of higher education in the state and to propose measures for further development. At the same time, since one of the most important key indicators for growth of Higher Education is GER, the government of India is targeting 50% GER in 2035. Drop-out rate and GER of schools in Mizoram at different stages are shown below in Pyramid P-1 and Pyramid P-2 respectively:

PYRAMID OF DROP OUT RATE			
Stages of Education	**2018-19**	**2019-20**	**Age group**
High School	10.64	20.35	14-15
Middle School	3.09	2.54	10-13
Primary School	7.52	7.8	6-9

Figure 1.5: Pyramid of Dropout rate

Source: udiseplus.gov.in

PYRAMID OF GER			
Stages of Education	**2018-19**	**2019-20**	**Age group**
Hr. Sec. School	52.13	53.7	16-17
High School	93.29	94.68	14-15
Middle School	100.26	102.94	10-13
Primary School	132.67	138.87	6-9

Figure 1.6: Pyramid of GER

Source: udiseplus.gov.in

1.2 Scope of the study

The purpose of this study is to ascertain the status of higher education in Mizoram and the various organizations that are either directly or indirectly involved in its growth. It also addresses a wide range of issues surrounding higher education including those involving students, teachers, and institutions as a whole. It is concerned with assessing the enrolment as well as gross enrolment ratio of higher education in Mizoram using statistics and machine learning technique. Higher education is an issue that is relevant not just to the field of education, but also to every academician of higher learning. In higher education, there are a plethora of approaches to choose from. The study of higher education can be approached differently by economists, social scientists, biologists, environmental scientists, statisticians, mathematicians, management experts, and so on. Because of this, a large number of works and studies in higher education can be classified as multidisciplinary. In today's world, multidisciplinary research is relatively prevalent and extremely fruitful for scholars. When it comes to achieving the goals of connecting and integrating several academic schools of thought, professions, or technology in the pursuit of a shared task, the term "collaborative innovation" is used. The current research endeavour falls in the category of interdisciplinary studies that span across mathematics and higher education institutions. Modern researchers find the use of mathematical models, statistical models, mathematical analysis, and statistical analysis to be quite productive. In the experimental sciences and social sciences, data analysis is a strong instrument to be used. For an in-depth data analysis to be accomplished, statistical methods such as sampling theory, computer technologies necessary for managing vast volumes of data, and expertise in analyzing information contained in various types of graphs are all required competencies. Meaningful results could be obtained from raw data after they have been analyzed and interpreted in a methodical manner. The expansion of multidisciplinary studies has had numerous beneficial effects across a wide range of fields of study, and in this context, examining the current situation of higher education in Mizoram is a hotly debated topic among academicians throughout the state.

1.3 Review of related literature

Education policy documents such as The Coleman Report (1966) and other major sources of information reveal contradictions on the issues of accomplishment and equality of educational opportunity across a wide range of social differences. The findings of (Johnes & Taylor, 1989) on the undergraduate non-completion rates of UK universities were that inter-university comparisons in the non-completion rate are of limited significance unless the disparities in scholastic ability of each university's intake students are taken into consideration. (Berdahl *et. al*., 1991) conducted a study in which they examined the quality and accessibility of higher education in the United Kingdom and the United States. (Epstein, 1992) examines the intricacy of the interplay between education and society in the United States in light of the paradoxes that exist therein. His book "Social Paradoxes of American Education" highlights a number of major paradoxes, including centralized control within a decentralized system, mass education versus quality education, equality of education and equality of educational opportunity, as well as paradoxes within culture-related issues such as the promotion of diversity within an assimilative system of education. Several versions of a generic model of quality assessment in higher education were proposed by (Vught & Westerheijden, 1994). (Keselman *et. al*., 1998) conducted a statistical analysis of educational researchers' practices. They discovered that researchers rarely check to see if validity assumptions are met, and as a result, they frequently adopt analyses that are not robust to assumptions being violated. It is the contention of (Feinberg & Soltis, 1998) that compulsory education ensures the replacement of outdated and dysfunctional habits, attitudes, and affiliations with newer and more functional ones. Although (Mugo, 1999) does not directly challenge the conceptions and theories of development that have been adopted by post-colonial Africa, he does point out that earlier conceptualizations and theories of development have existed in traditional African societies, which he considers to be important in a nation's development. Women in higher education were investigated by (Stake & Hoffmann, 2001), who looked at how social attitudes, activism, and personal confidence changed as a result of their presence. (DesJardins *et. al*., 2003) developed a model of the application, admission, financial aid determination, and enrolment decision processes that was jointly recommended.

(Riggert *et. al*., 2006) reviewed the empirical literature and examined the relationship between student employment and higher education. They proposed possible explanations for inconsistencies, such as the challenges posed by methodological issues and the lack of theoretical conceptualization. Several years ago, (Hristova & Zhelezarov, 2006) established a model of criteria for measuring and assessing the quality of university education. This model is still in use today. Using data from the European higher education region, (Schwarz & Westerheijden, 2007) conducted a comparative analysis on accreditation in the context of valuation activities in higher education. (Sanusi & Oyama, 2008) used statistical data analysis to investigate the Japanese government's subsidy policy for private universities. They discovered that the number of faculty members has the greatest influence on general subsidies, while the number of students has the greatest influence on special subsidies. Online distance learning was researched by (Puthe, 2008), who concluded that the advantages of online distance learning include access to information 24 hours a day, the use of up-to-date content materials, self-paced learning, tailored courses, and cost effectiveness. (Goyal, *et. al*., 2011) investigated the impact of student satisfaction and the usability of the Internet on academic achievement.

(Said, 2011) created a pedagogical model connected to new teaching-learning opportunities in the classrooms of Colombian institutions in order to support the use of information and communications technology (ICT) in education. As a result of their research, (Bastedo & Jaquette, 2011) discovered that low-income students had achieved significant gains in their academic course results during the 1970s, as well as the dynamics of higher education stratification. (Perna *et. al*., 2014) investigated the foreign scholarship program in higher education and used Stata statistical software to analyze the data. According to (Alam *et. al*., 2014), a case study of affiliates under the national university of Bangladesh was conducted in order to determine quality assurance techniques for affiliated institutions of higher learning. Their research revealed that politicians are blind to the notion that the quality of education in the affiliated institutions is essential. (Reale, 2014) investigated the use of quantitative techniques in comparative analyses specifically in the field of higher education. A key methodological problem for accurate international comparisons in higher education studies was addressed, as were the difficulties encountered when using measurements in comparative studies. Concerns about issues that need to be addressed in order to improve methodological robustness as well as the possibility of using quantitative tools in comparative studies were also discussed. A recent study by (Zmas, 2015) examined how the financial crisis has influenced higher education policies in Greece, and he concluded that the imposition of relevant policies has broader roots that can be traced back to the contradictions that have been observed during the democratization and modernization of universities over the last 40 years. (Strauss & Borenstein, 2015) suggested a system dynamics model for long-term planning of undergraduate education in Brazil, which was adopted by the Brazilian government. A system dynamics model for analyzing long-term policies involving undergraduate programs in Brazil had been developed and applied by them, and they had presented their findings in this paper.

(Prakash, 2007) explored the growth and financing of higher education in India as well as the implications of these trends. (Barbhuiya, 2014) conducted research on the economics of higher education financing in Mizoram. The statistical methods used by (Rupon, 2012) to analyze school dropout across Indian states and union territories came to the conclusion that school dropout is caused by a variety of factors and that appropriate initiatives such as poverty alleviation and school infrastructure improvement, increased numbers of trained teachers, and curriculum adaptation to meet current needs would be required for mitigating the problem. Using multiple data sources, the authors (Gupta & Gupta, 2012) presented and analysed the development and current situation of higher education in India. They also identified the most significant difficulties that India's higher education sector is experiencing. The situation of science education in Mizoram was examined in depth in a critical study by (Zohmingliani, 2012). (Kareena & Manoj, 2013) conducted comparative research of components of value-based higher education systems in six nations—the United Kingdom, China, the United States, Australia, Brazil, and South Africa—in comparison to the Indian higher education system. Education reforms were recommended, and the key components of controlling and delivering greater value to students in India's higher education system were explained in detail. (Ramesh, 2013) conducted an investigation into Indian higher education and discovered that the sector has seen a tremendous increase in its institutional capacity in the years since Independence, but that the challenges of higher education have been caused by low college enrolment, an unemployment crisis among unskilled labour, and a lack of flexibility in the education sector. (Singh, 2013) conducted a study on the issues and opportunities of higher education in Manipur and concluded that preserving and improving higher education in Manipur is essential for the state's and nation's bright futures to become one of the superpower countries in the not-too-distant future.

On the basis of their paper, "Higher Education Scenario of the North-Eastern India," (Konwar & Chakraborty, 2013) concluded that the North-Eastern region of India is lagging behind in terms of quality education and has a dearth of constructive higher educational institutions when compared to other regions of India. According to (Gaikwad & Solunke, 2013), higher education in India is growing at a rapid pace. They argue that universities and colleges should provide sufficient employable skills to ensure that students' employability improves. (Bhagat, 2013) researched the expansion of higher education in Jammu and Kashmir and the consequences of this expansion. He makes ten recommendations, one of which is that students be encouraged to participate in research activities. On the basis of their paper "A statistical study on higher educational institutions in India," (Neelaveni & Manimaran, 2014) conclude that the learning environment should be improved to meet the requirements of modern media and that the higher education department should provide favourable financial resources to follow updated teaching methods and also restructure modern infrastructural requirements. The problems and potential of higher education in India were investigated by (Agrawal, 2014), who concluded that the development of higher education has come at the expense of quality. From the pre-globalization era to the post-globalization era, (Kaur, 2014) examined the evolution of higher education in India. The authors of the research paper, "Excellence through Equity: Challenges before Higher Education in India" (Misra & Bal, 2014), argue that the entire Indian school system, including primary and secondary education, vocational education, and higher education, needs to be restructured.

With the help of a nationally representative household survey and a multinomial logistic regression approach, (Chea, 2019) provides empirical evidence on the trends of how demand-side factors predict the probabilities of enrolment in higher education in

Cambodia between 2004 and 2014. He asks the question "Does higher education expansion in Cambodia make access to education more equal?" According to the findings, Cambodia's expansion of higher education appears to be skewed in favour of students from wealthy families who live in the country's capital. Students from underprivileged families, on the other hand, are more likely to profit from the expansion of higher education, provided that they are able to complete high school. (Lu & Zhang, 2019) conducted a study on China's higher education expansion and rural children's involvement in senior high school, and their findings indicate that the opportunity to attend college plays a significant role in rural children's schooling decisions. They discovered that in order to overhaul China's higher education system, more educational resources should be allocated to rural areas, according to their research. They go on to say that rural children whose parents work in cities should be given greater opportunity to attend urban schools during the urbanization process; otherwise, it will be very difficult to achieve universal access to higher education among Chinese children. (Meier & Schiopu, 2020) investigated the expansion of enrolment and the level of differentiation across higher education institutions in the United States. The researchers discovered that, whereas the impact of a higher college wage premium on enrolment growth is widely established, the link between institution quality differentiation and student outcomes in this setting has gotten less attention. In order to address this issue, they presented alternative models of higher education systems. A differentiated education system, such as the one used in the United States, can lead to lower academic standards as a low-quality segment emerges. However, a uniform standard, such as the one used in most European countries, can lead to lower academic standards as the interests of marginal students are taken into consideration when determining the standard. In the absence of complete knowledge regarding graduates' abilities, employers place greater emphasis on the reputation of the university rather than on the individual human capital signal. As a result, pupils with medium abilities may exert less effort and demonstrate less proficiency when their distinctiveness is increased. As a result, they argue that the trend toward more unequal societies increases political support for systems with sharply differential educational opportunities for high- and low-ability kids, respectively.

(Agasisti & Bertoletti, 2020) evaluated the impact of regional higher education systems (HESs) on economic growth using data from 284 European regions (NUTS 2) over an 18-year period, and their findings were published in the journal Higher Education (from 2000 to 2017). With indicators on university concentration, HES size and performance, as well as other significant characteristics, the empirical framework precisely predicts the heterogeneity of high-performing schools (HESs). A novel and integrated dataset was used in this study, which was produced by gathering and merging indications from a variety of various data sources (Eurostat, OECD, WHED, and InCites). According to the findings, an increase in the number of institutions in a region is associated with greater economic growth in that region. The quality of research and a specialization in STEM disciplines are the key drivers of universities' favourable impact on the economic development of their respective regions. A brief history of this phenomenon as well as its current status were researched and highlighted by (Uralov, 2020), who also discussed the rationales for the globalization of higher education in Uzbekistan. (Wild & Heuling, 2020) investigated student dropout and retention. They examined a sample of 19,769 cooperative students in Germany, which included 852 dropouts, as part of their research. In order to identify markers of dropout risk, a Cox regression analysis was carried out. The findings indicate that, in addition to the recognized predictors of cognitive abilities, one's personality and devotion to one's firm are important. (Yang *et. al*., 2020) used time-series analysis to forecast student enrolment and teacher statistics. In this study, combining a whale optimization algorithm (WOA) and support vector regression (WOASVR), it was proposed to determine trends of student and teacher numbers in Taiwan for higher accuracy in time-series forecasting analysis. To select the most suitable support vector kernel parameters, WOA was applied. Data collected from the Ministry of Education's datasets of student and teacher numbers between 1991 and 2018 were used to examine the proposed method. Analysis revealed that the numbers of students and teachers decreased annually except in private primary schools. A comparison of the forecasting results obtained from WOASVR and other common models indicated that WOASVR provided the lowest mean absolute percentage error (MAPE) and root mean square error (RMSE) for all analyzed datasets.

1.4 Rationale of the study

Enrolment is one of the most significant indicators of higher education. For a better knowledge of higher education as well as the creation and application of the policy, a thorough investigation of enrolment-related issues is essential. Despite this, Mizoram in particular and all of India have relatively little research on enrolment in higher education. The majority of research focuses on GER of higher education, but the current study also uses statistical and machine learning technique to provide systematic and scientific knowledge of the problem and predict potential outcomes. The study's conclusions would be useful for developing and implementing policy.

In view of the above, the current study, titled **"Mathematical and Statistical Analysis of Higher Education in Mizoram,"** will be beneficial for Mizoram in particular in comprehending higher education in general.

1.5 Objectives of the study

The thesis has the following objectives:

- to investigate student dropout at various stages in school.
- study the problems of low GER in higher education and propose measures to increase it.
- to develop a mathematical model of GER.

1.6 Methodology

Relevant data and information from secondary sources, including the UNESCO report, the AISHE annual report published by the Ministry of Education, Government of India, and data from the state government, among others, were gathered in order to investigate the current state of higher education in Mizoram. For school education, data were also gathered through the National Sample Survey Organization (NSSO) and the Unified District Information System for Education (UDISE). Statistical analysis and machine learning methods were employed to analyze and evaluate the crucial data that had been collected for the present study. The enrolment issue was the focus of our study, and we had suggested solutions to increase GER. Additionally, the performance of LSTM (machine learning methodology) was shown to be superior to ARIMA (statistical approaches) when used to analyze GER of higher education data. Five types of machine learning, in specific to regression modeling were also compared to analyze the enrolment of higher education in Mizoram. In this comparison, Bagging regression is found to be superior among the candidates and is used to forecast future enrolment.

1.7 Chapterization

Chapter 1: Introduction: The first chapter introduces the topic of the study by outlining the current state of higher education. The chapter also includes a review of the relevant literature and the study's rationale, objectives, and methodology.

Chapter 2: Student dropout at different levels of school education and its implication on higher education GER: The chapter analyzes UDISE and NSSO data to identify drop-outs at various levels of education and by demographics. It also investigates several aspects of dropout in Mizoram. In order to study the effect of student dropout at different levels of school education, GER of higher education is calculated considering there were no dropouts at different levels of school. Future prediction was also done in this regard.

Chapter 3: GER of higher education in Mizoram: The chapter offers a comprehensive examination of higher education enrolment and GER in Mizoram. The chapter also studies the relationship between the GER of higher education and the GER of different levels of school education.

Chapter 4: Data analysis and prediction of GER using machine learning and statistical methods: Two cutting-edge forecasting methodologies, ARIMA and LSTM, are compared in this chapter. By converting the yearly enrolment dataset into the comparable GER, time series data were pre-processed using this method. For future GER forecasting, the LSTM machine learning model was selected during the model selection phase. As a result, the study's second phase is aimed at carrying out a more thorough performance examination of various machine learning architectures.

Chapter 5: Conclusion: The chapter summarizes the thesis and highlights key findings of the study.

CHAPTER 2

STUDENT DROPOUT AT DIFFERENT LEVELS OF SCHOOL EDUCATION IN MIZORAM AND ITS IMPLICATION ON HIGHER EDUCATION GER

2.0 Introduction

One of the fundamental conditions for the growth of humanity is education. It serves as the cornerstone for both personal and societal growth. Employment options are expanded, and income levels are raised with schooling. A country's literacy rate is used to gauge its level of education. People's wellbeing and progress are both seriously hampered by illiteracy. Despite various government initiatives, the major issue of school dropout persists in many countries. Dropping out results in the waste of all the resources the government, society, community, and family have invested, as well as the labour and wealth of the entire country, with negative long-term consequences.

One of the main causes of India's poor literacy rate is the issue of dropouts, since these children do not get the education they need. Although the causes may change at different times, India's dropout problem affects students of all ages and is present not only in schools but also at all levels of the education system, starting from primary school to higher education. Reasons for dropping out vary from adverse financial conditions to personal problems. A dropout can be defined as an "ever-enrolled person" who does not complete the last level of education for which he or she has enrolled and is currently not attending any educational institution (Bote-Lorenzo & Gómez-Sánchez, 2017). According to the UDISE report 2021-2022 (https://dashboard.udiseplus.gov.in/) by the Ministry of Education, India, the dropout rates at the primary, upper primary, and secondary levels among boys and girls are (1.55, 1.33), (2.74, 3.31), and (12.96, 12.25), respectively. Despite the fact that the Right to Education Act (RTE) (2009) and Sarva Shiksha Abhiyan (2010) have significantly reduced the dropout rate in India, a sizable portion of children still do not finish their schooling. There have been many empirical and theoretical studies done that look at the dropout problem from different points of view.

(Chugh, 2011) noted that family background (lack of money and lack of desire for education) are significant factors in why children living in Delhi's slum regions drop out of school. She added that the dropout issue worsens in the lower and upper secondary levels. Research was done on the state of the educational system in relation to gender disparity (Amirtham & Kundupuzhakkal, 2013). They discovered that early marriage, a rigid social structure, and a lack of facilities suited for female students all contribute to the dropout problem being more severe among girl children. According to (Gouda & Sekher, 2014) research, children from Muslim, Scheduled Caste, and Scheduled Tribe families had a higher dropout rate than other children. They also noticed that dropout rates among children of illiterate parents were four times higher than those of literate parents and that the likelihood of dropout among children of parents who were not working was rather high. In the backdrop of the absence of facilities appropriate for female students, (Prakash *et. al*., 2017) provided a vivid picture using examples from Karnataka villages where the vulnerability of adolescent girl children is another cause for leaving school. (Marphatia *et. al*., 2019) developed a biosocial life-course conceptual approach to investigate maternal and household predictors of secondary school dropout, and to ascertain whether the consequences of dropout differ between girls and boys. They analysed longitudinal biomedical data on 648 mother-child dyads from rural Maharashtra, India and found that both maternal (low education, early marriage age, shorter pregnancy duration) and household (low paternal education, low socio-economic status) traits independently predicted dropout.

(Weybright *et. al*., 2017) used survival analysis to identify the risk of dropping out of secondary school for male and female adolescents and found that being male, not living with one’s mother, smoking cigarettes in the past month, and lower levels of leisure-related intrinsic motivation significantly predicted dropout. (Mussida *et. al*., 2019) studied the impact of secondary school dropout on the work outcomes of young people in ten developing countries. They found that secondary school dropout decreases the probability of being employed in non-elementary occupations, suggesting that unskilled workers fail to meet the increasing demand for a skilled workforce. (Sharma & Levinson, 2019) studied the association between travel time to the lower secondary and secondary public schools of Nepal and the dropout grade before leaving secondary school. Using an ordered logit model, they found that as the travel time to the school increases, students were more likely to dropout from the school system in earlier grades.

Adolescence is a time for developing new skills. It is not only a period of opportunity but also a time of vulnerability to risky behaviour, which can have long-term effects, particularly on one's health, profession, and education. In India, over a fifth of children between the ages of 6 and 11 are not in school, followed by a third of teenagers between the ages of 12 and 14. This is according to a joint study by the UNESCO Institute for Statistics and Global Education Monitoring. According to (Majumder & Mitra, 2020), children from wealthy families with educated parents have a lower rate of dropping out in both the rural and urban sectors. Large families and having many younger siblings, however, have a detrimental impact on school participation in the rural sector, but similar factors have no impact on urban children.

Student dropouts at various levels have a significant impact on enrolment status (Hone & El Said, 2016). Domestic work, economic conditions, and a lack of interest were discovered to be the most common reasons for dropping out of school. About 30.2% of the girls gave domestic work as the reason for discontinuing education, and about 36.90% of boys left their studies to support their families. It becomes challenging for girls to continue studying because of concerns about their safety (Romero & Ventura, 2020). They face sanitary problems due to poor school facilities, ultimately forcing them to stay home. Considered a liability, many girls are forced to stay at home or are forced to get married at an early age (13.2%). Many children believe there is no point in studying if they have to do the same job as their parents. As a result, they drop out of primary school (NSS Report No.585, 2018). More than 30% of the children involved in the survey showed a lack of interest in studies; they preferred to drop out because whatever was being taught in schools barely intrigued them (Kizilcec *et. al*., 2013). India is also dealing with the problems of inclusion and equality; children from marginalized sections or with physical disabilities or health issues must leave schools when they face hostile behaviour from their peers.

In this section, descriptive statistics were used to identify the causes of student dropouts in Mizoram using data from the NSSO's 75th round. The GER for higher education is calculated in such a way that the student dropout rate throughout all levels of education is zero. Higher education GER prediction is carried out in this case. The goal of this study is to determine how student dropout affects GER. In brief, one of the key measures of the status of academic enrolment is the GER (Dalipi, *et. al*., 2018). At the state level, prediction of future GER has been done using statistical techniques. Linear and non-linear models are fitted to time-series data to obtain aggregate enrolment in higher education. This study is carried out by assuming that state enrolment is not affected by dropouts, or, in other words, by neglecting student dropouts.

2.1 Growth of school education in Mizoram

Mizoram is a small state in northeastern India, covering a total land area of 21,087 square kilometers. It is located between the longitudes of 92.15° and 93.29° E and the latitudes of 21.58° and 24.35° N. As per the 2011 census, Mizoram has a population of over 10.91 lakh. Out of the total population, about 5.52 lakh are male and 5.38 lakh are female. The state's population forms only 0.09 percent of the total population of India. Mizoram is a highly literate state as per the 2011 census; the literacy rate is 91.58 percent. This is far higher than the national literacy rate of 74.04 percent. Mizoram has the second highest literacy rate in India, trailing only Kerala (Chhuanawma, 2015). In Mizoram, education began with the efforts of Christian missionaries. The Mizo people did not have their own script until the advent of the pioneer missionaries. The two missionaries F.W. Savidge and J.H. Lorrain prepared a

Roman script (A, AW, and B) for Lushai which is still used today. Schools were at the initial stage of being opened with the efforts of the missionaries, and the first primary school was started on 2nd April, 1894 (Kumar, 1994). Education grew slowly because the government at the time did not prioritize education development, particularly beyond elementary school. After a long wait, the first high school opened in 1944.

School education in Mizoram is categorized broadly as elementary and secondary. Elementary education covers classes 1 to 8, and secondary education covers classes 9 to 12. In Mizoram, the School Education Department looks after government, aided, and private schools except the Central Schools and the elementary schools within the three district councils of Lai Autonomous District Council, Mara Autonomous District Council, and Chakma Autonomous District Council. The growth of schools in Mizoram is shown in the following Table 2.1:

Table 2.1: Growth of schools in Mizoram

Year	Primary Schools	Middle Schools	High Schools	Higher Secondary Schools	Grand total
1894	1	Nil	Nil	Nil	**1**
1903	3	Nil	Nil	Nil	**3**
1947	258	22	2	Nil	**282**
1972	425	184	70	Nil	**961**
1986	1017	443	154	Nil	**1614**
1996	1263	702	300	16	**2281**
1997	1318	733	-	-	**2396**
1998	1248	-	-	-	**1608**
1999	1226	748	383	18	**2344**
2000	1224	734	370	30	**2371**
2001	1377	851	409	33	**2631**

2002	1504	911	-	47	**2871**
2004	1481	939	448	67	**2935**
2005	1688	1121	474	125	**3408**
2006	1618	1081	492	80	**3271**
2007	1668	1090	496	82	**3336**
2008	1761	1117	342	59	**3279**
2009	1763	1180	362	70	**3375**
2010	1821	1353	538	98	**3810**
2011	1855	1383	543	113	**3894**
2012	1831	1381	584	118	**3914**
2013	1876	1408	612	127	**4023**
2014	1946	1514	610	132	**4202**
2015	1950	1511	614	138	**4213**
2016	1968	1542	640	163	**4313**
2017	1969	1580	669	175	**4393**
2018	1956	1553	691	186	**4386**
2019	1955	1550	707	197	**4409**

Source: Statistical Wing, Directorate of School Education; Government of Mizoram as on 16th June 2022

2.2 GER of School education in Mizoram

Because the missionaries and the government did not work together to develop higher education, it did not begin until 1958. To study higher education enrolment, one must first study school enrolment, because only students who have passed out of high school can be admitted to higher education. Enrolment in school education will be highlighted, and the trend of school education enrolment will be studied from the lower level to the higher level. The GER of school education at different levels of Mizoram during the last ten years is shown in Table 2.2.

Table 2.2: GER of school education in Mizoram (2012-13 to 2021-22)

Academic Level	Category	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20	2020-21	2021-22
Primary (I-V)	Girls	147.49	126.34	122.9	123.61	116.42	123.77	131.68	138.11	145.38	159.22
	Boys	154.46	130.79	127.88	128.24	121.28	127.14	133.62	139.59	146.65	158.58
	Overall	151.03	128.6	125.43	125.96	118.89	125.49	132.67	138.87	146.03	158.89
Upper Primary (VI-VIII)	Girls	114.43	100.09	100.96	101.19	96.62	97.39	100.22	103.11	106.29	110.81
	Boys	116.74	102.45	103.76	104.66	99.1	98.69	100.3	102.78	104.79	108.33
	Overall	115.61	101.29	102.39	102.96	97.88	98.06	100.26	102.94	105.52	109.54
Secondary (IX-X)	Girls	90.45	92.24	92.9	95.21	84.18	87.3	94.54	98.43	97.23	97.86
	Boys	88.19	90.85	90.5	91.53	79.73	81.42	92.11	91.1	86.18	89.05
	Overall	89.3	91.53	91.67	93.32	81.9	84.29	93.29	94.68	91.58	93.36
Higher Secondary (XI-XII)	Girls	44.06	47.42	48.68	50.73	49.73	49.07	55.18	56.34	56.89	65.69
	Boys	43.94	45.33	48	49.17	51.16	43.93	49.17	51.16	51.33	57.09
	Overall	44	46.36	48.34	52.13	53.7	46.46	52.13	53.7	54.06	61.3

It can be seen from the above Table 2.2 that overall GER is highest at the primary school level, ranging from 118.89 to 158.89 during the last ten years, and slowly decreasing upward in the upper primary school level, ranging from 97.88 to 115.61, and again further decreasing in the secondary school level, ranging from 81.9 to 94.68, and drastically decreasing in the higher secondary school level, in the range of 44 to 61.3. The GER pattern is consistent across all three categories: girls, boys, and overall.

Table 2.3: The mean GER for different levels of school education

	Variables	Primary	Upper Primary	Secondary	Higher Secondary
Girls	GER (±SD)	133.49 (±13.62)	103.11 (±5.76)	93.03 (±4.67)	52.38 (±6.23)
Boys	GER (±SD)	136.82 (±12.58)	104.16 (±5.29)	88.07 (±4.339)	49.03 (±4.03)
Overall	GER (±SD)	135.186 (±13.05)	103.65 (±5.43)	90.49 (±4.20)	51.22 (±5.06)

From the above Table 2.3, it is clearly evident that GER is rapidly decreasing from the secondary level to the higher secondary level. Hence, GER at the secondary and higher secondary levels will be the main focus of the present study. To be more precise, the GER of school education at the secondary (high school) and higher secondary (higher secondary school) level will be calculated and analyzed district-wise in the following tables:

Table 2.4: GER of High School (2015-16 to 2019-20)

	2015-16			2016-17			2017-18			2018-19			2019-20		
	F	M	O	F	M	O	F	M	O	F	M	O	F	M	O
Aizawl	111.2	103.1	107.1	99.6	90.7	95.1	102.7	91.8	97.2	106.8	103.3	105	113.1	102.6	107.8
Champhai	102.5	91.2	96.7	92	82.4	87.1	94.1	85.7	89.8	95.7	89.8	92.6	101.5	89.8	95.5
Kolasib	85.4	85.1	85.2	75.7	76.4	76	82.5	76	79.2	90.1	86.1	88	94.4	89.7	92
Lawngtlai	67.9	77.7	73	63	78.4	71	59.6	72.3	66.2	79.8	94.1	87.1	82.9	94	88.6
Lunglei	92.8	93.9	93.4	73.8	67.6	70.6	78.6	76.8	77.7	94.3	90.5	92.3	88.7	82.5	85.5
Mamit	71.3	69.9	70.6	62.5	57.9	60	69	63.6	66.2	67.4	64.8	66	72.9	66.4	69.5
Saiha	90.2	81.8	85.8	82.3	69.2	75.4	87.2	69.4	77.9	93.2	81.8	87.2	97.9	80.3	88.7
Serchhip	94.4	96	95.2	86.5	93.2	89.8	88.7	88.3	88.5	91.6	90.9	91.3	99.7	93.3	96.5
Mizoram	**95.2**	**91.5**	**93.3**	**84.2**	**79.7**	**81.9**	**87.3**	**81.4**	**84.3**	**94.5**	**92.1**	**93.3**	**98.4**	**91.1**	**94.7**

Table 2.5: GER of Higher Secondary School (2015-16 to 2019-20)

	2015-16			2016-17			2017-18			2018-19			2019-20		
	F	M	O	F	M	O	F	M	O	F	M	O	F	M	O
Aizawl	79.6	82.6	75.0	75.5	70.1	72.8	75.2	69.1	72.2	81.6	75.9	78.8	85.7	79.6	82.6
Champhai	30.7	35.4	31.8	34.9	28.9	31.8	32.9	26.5	29.6	38.0	29.3	33.6	40.4	30.7	35.4
Kolasib	36.0	37.3	32.6	35.7	32.1	33.9	33.2	33.1	33.1	37.7	31.3	34.5	38.7	36.0	37.3
Lawngtlai	39.4	35.6	26.5	28.6	36.8	33.0	21.4	28.2	25.0	27.4	37.1	32.5	31.4	39.4	35.6
Lunglei	44.7	48.2	47.4	39.9	35.3	37.5	42.9	34.4	38.5	52.5	42.5	47.4	51.8	44.7	48.2
Mamit	10.8	11.2	8.4	9.1	9.7	9.4	10.8	8.6	9.6	13.4	11.1	12.2	11.7	10.8	11.2
Saiha	34.7	36.2	46.1	49.0	43.0	46.0	47.9	45.1	46.5	53.1	45.1	49.1	37.7	34.7	36.2
Serchhip	40.5	44.1	75.0	40.9	34.0	37.4	42.2	33.6	37.8	48.2	37.7	42.8	47.9	40.5	44.1
Mizoram	**51.2**	**53.7**	**31.8**	**49.7**	**45.5**	**47.6**	**49.1**	**43.9**	**46.5**	**55.2**	**49.2**	**52.1**	**56.3**	**51.2**	**53.7**

Enrolment data for high school and higher secondary school in all districts has been generated from UDISE (udise.gov.in), and the age population of all districts for various years has been estimated based on 2011 Census data. The GER of high school and higher secondary school for all districts was then calculated for five years, as shown above in Tables 2.4 and 2.5. Looking at the female, male, and overall GER of high schools from 2015-16 to 2019-20, Aizawl district has the highest figure in all three categories throughout the time period. Mamit and Lawngtlai districts lagged behind in the first three years. An improvement is seen in the last two years for Lawngtlai district, whereas Mamit district is still lagging behind even in the last two years.

2.3 Dropout rate of school education in Mizoram

Table 2.6: Dropout rate of school education in Mizoram (2013-14 to 2021-22)

		Primary			Upper Primary			Secondary		
	Soc. Cat.	Girls	Boys	Over-all	Girls	Boys	Over-all	Girls	Boys	Over-all
2013-14	Gen.	0	0	0	0	0	0	8.38	0	1.9
	SC	0	0	0	0	0	0	0	0	0
	ST	24.51	24.82	24.67	19.85	20.05	19.95	22.47	24.17	23.33
	OBC	0	0	0	0	0	0	7.14	15	11.21
	Total	23.94	24.26	24.11	19.22	19.35	19.29	20.65	22.18	21.42
2014-15	Gen.	0	0	0	0	0	0	20.11	3.61	11.3
	SC	0	0	0	0	0	0	0	0	0
	ST	13.52	12.71	13.1	5.57	6.93	6.27	20.05	22.55	21.31
	OBC	80.85	79.25	79.98	63.74	58.93	61.36	25	33.93	29.63
	Total	13.38	12.57	12.97	5.38	6.61	6.02	17.37	20.01	18.7
2015-16	Gen.	0	0	0	0	0	0	17.77	28.29	23.66
	SC	0	0	0	0	0	0	14.2	0	6.01
	ST	10.7	10.84	10.77	4.96	6.39	5.7	20.88	23.25	22.07
	OBC	0	0	0	0	0	0	0	0	0
	Total	10.03	10.17	10.11	4.06	5.45	4.78	20.73	23.02	21.88

2016 -17	Gen.	0	0	0	0	0	0	13.07	17.45	15.21
	SC	68.08	62.6	65.12	60.83	62.34	61.62	87.69	84.87	86.2
	ST	15.23	15	15.11	8.62	10.17	9.43	27.74	31.32	29.54
	OBC	0	0	0	0	0	0	0	0	0
	Total	15.49	15.24	15.36	9.06	10.63	9.88	28.87	32.46	30.67
2017 -18	Gen.	40.42	40.17	40.28	3.12	18.5	11.16	0	0	0
	SC	0	0	0	0	0	0	0	0	0
	ST	6.93	8.01	7.48	5.97	8.36	7.2	19.25	24.73	21.98
	OBC	0	0	0	0	0	0	0	0	0
	Total	7.41	8.56	8.01	5.63	8.22	6.97	16.87	22.61	19.73
2018 -19	Gen.	27.54	25.82	26.62	0	0	0	65.16	58.88	62.24
	SC	0	0	0	0	0	0	57.79	63.09	60.48
	ST	7.28	8.21	7.75	3.06	4.75	3.92	7.72	9.08	8.4
	OBC	0	0	0	0	0	0	0	0	0
	Total	7.05	7.96	7.52	2.3	3.83	3.09	10.13	11.17	10.64
2019 -20	Gen.	12.31	13.35	12.87	14.89	19.74	17.6	43.35	43.3	43.32
	SC	0	0	0	11.79	4.52	8.08	13.73	13.66	13.71
	ST	7.6	7.96	7.78	0.88	3.53	2.23	17.2	22.99	20.12
	OBC	60.2	56.36	58.19	49.02	30.53	40.1	0	0	0
	Total	7.58	8.01	7.8	1.24	3.79	2.54	17.47	23.18	20.35
2020 -21	Gen.	4.04	5.6	4.85	0	0	0	27.5	23.23	25.37
	SC	2.16	0	0	0	3.21	0	52.91	46.7	49.87
	ST	7.8	8.65	8.24	3.9	7.29	5.63	17.44	22.03	19.69
	OBC	0	0	0	14.29	1.69	7.83	0	5.88	0
	Total	7.63	8.46	8.06	3.46	7.1	5.32	17.82	22.24	19.99
2021 -22	Gen.	12.32	32.31	22.55	21.43	26.11	23.87	42.31	41.67	41.97
	SC	0	0	0	1.83	0	0	42.75	31.39	37.09
	ST	5.46	6.8	6.16	1.36	3.62	2.51	10.37	12.68	11.48
	OBC	57.79	54.42	56.15	35.56	16.42	24.11	0	0	0
	Total	5.58	7.08	6.35	1.64	3.78	2.73	10.83	13.06	11.9

Mizoram's school dropout rate is lowest at the primary level, rising at the upper primary level, and rising further at the secondary level between 2013-14 and 2021-22. Therefore, the main focus of the present study will be on the secondary level, as the dropout rate is highest at this level. Secondary school dropout in Mizoram is analysed district-wise and category-wise in the following tables.

2.4 Secondary School Dropout of Mizoram: District-wise and Category-wise

Table 2.7: Mean dropout rate of Secondary School (Girls, 2013-14 to 2021-22)

District	Social Category	General	OBC	SC	ST	Overall
Aizawl	Mean dropout	27.04	10.03	27.27	8.01	7.02
	(±SD)	25.72	17.40	30.50	5.51	5.00
Champhai	Mean dropout	0	0	10	31.07	31.15
	(±SD)	0	0	25.87	9.68	9.56
Kolasib	Mean dropout	23.46	7	23.64	22.78	23.58
	(±SD)	14.33	18.11	33.38	6.20	5.54
Lawngtlai	Mean dropout	12.79	0	0	29.00	28.81
	(±SD)	24.50	0	0	13.10	12.99
Lunglei	Mean dropout	39.20	5.56	34.11	23.12	24.46
	(±SD)	28.02	14.37	33.87	8.93	9.68
Mamit	Mean dropout	0	0	0	41.11	41.11
	(±SD)	0	0	0	5.36	5.46
Saiha	Mean dropout	22.96	0	10.39	21.67	21.75
	(±SD)	31.86	0	26.89	7.64	7.56
Serchhip	Mean dropout	17.59	0	12.5	23.67	23.61
	(±SD)	25.15	0	24.83	5.33	5.38

Table 2.8: Mean dropout rate of Secondary School (Boys, 2013-14 to 2021-22)

District	Social Category	General	OBC	SC	ST	Overall
Aizawl	Mean dropout	24.42	9.89	21.39	9.56	8.66
	(±SD)	26.42	17.73	31.57	6.47	6.04
Champhai	Mean dropout	0.00	0.00	12.37	34.83	34.92
	(±SD)	0.00	0.00	27.78	9.06	8.84
Kolasib	Mean dropout	25.45	18.11	27.26	26.94	27.59
	(±SD)	17.88	31.64	38.56	6.60	5.28
Lawngtlai	Mean dropout	17.54	0.00	7.41	28.43	28.36
	(±SD)	30.46	0.00	19.17	9.54	9.74
Lunglei	Mean dropout	31.01	7.41	33.83	27.34	28.65
	(±SD)	34.79	19.18	35.23	13.36	14.10
Mamit	Mean dropout	0.00	5.56	0.00	43.54	43.44
	(±SD)	0.00	14.37	0.00	6.43	6.56
Saiha	Mean dropout	22.78	0.00	20.23	24.87	24.95
	(±SD)	39.60	0.00	44.46	10.09	8.80
Serchhip	Mean dropout	33.49	12.59	5.56	29.12	29.11
	(±SD)	29.33	24.01	22.27	5.76	6.08

Table 2.9: Mean dropout rate of Secondary School (Overall, 2013-14 to 2021-22)

District	Social Category	General	OBC	SC	ST	Overall
Aizawl	Mean dropout	25.60	9.93	24.21	8.76	7.82
	(±SD)	25.57	17.54	30.09	6.27	5.74
Champhai	Mean dropout	0.00	0.00	11.13	32.95	33.02
	(±SD)	0.00	0.00	26.61	9.00	8.86

Kolasib	Mean dropout	24.19	17.97	25.39	24.82	25.60
	(±SD)	14.51	31.67	36.71	6.32	5.17
Lawngtlai	Mean dropout	15.94	0.00	8.89	28.71	28.58
	(±SD)	27.82	0.00	23.00	11.68	11.77
Lunglei	Mean dropout	30.70	15.90	33.21	25.21	26.07
	(±SD)	33.52	28.28	34.32	10.78	11.38
Mamit	Mean dropout	0.00	5.56	0.00	42.42	42.39
	(±SD)	0.00	14.37	0.00	5.23	5.32
Saiha	Mean dropout	26.93	0.00	21.20	23.23	23.30
	(±SD)	33.57	0.00	41.40	9.25	7.99
Serchhip	Mean dropout	28.59	16.19	11.73	26.28	26.28
	(±SD)	28.29	28.63	24.76	4.61	4.73

Among the districts, Mamit has the highest female dropout rate at the secondary level, whereas Aizawl has the lowest overall dropout rate at the secondary level throughout the period. The mean female dropout rates of different districts range from 7.02 in Aizawl to 41.10 in Mamit. Apart from this, Saiha, Serchhip, Kolasib, Lunglei, and Lawngtlai districts fall in the range of 20 and 30, whereas Champhai district is in the range of 30 and 40.

The mean male dropout rate for female in different districts ranges from 8.66 in Aizawl to 43.44 in Mamit. There are five districts with rates falling between 20 and 30 (Saiha, Kolabib, Lawngtlai, Lunglei and Serchhip) and the mean male dropout rate of Champhai lies between 30 and 40.

The lowest and highest mean overall dropout rates are in Aizawl and Mamit, at 87.82 and 42.39, respectively. Moreover, there are five districts with rates falling between 20 and 30 and one district between 30 and 40. Since the majority of Mizoram's population fall under the Scheduled Tribe category and those belonging to other categories are relatively small in number, which may be negligible for the whole state, the dropout rates of different categories of General, Other Backward Class, Scheduled Caste and Scheduled Tribes have not been explored deeply.

From the detailed study of the GER of school education in Mizoram, it is clear that student enrolment decreases from the primary level to the higher secondary level. The trend also shows that GER is drastically decreasing from high school level to higher secondary level, which indicates that there are issues to be addressed in the level between the high school and higher secondary levels. District-wise analysis of school education GER at the secondary level in Mizoram shows that Mamit district is far lagging behind other districts. Other districts like Champhai and Lawngtlai are also at low levels, while Aizawl district, which comprises the state capital, is doing well on the other hand.

Similarly, a detailed examination of dropout rates in the Mizoram school education system reveals that dropout rates are increasing from the primary to the secondary levels. From the student dropout trend in school education, it is clear that the dropout rate jumps at the level of secondary school, calculated for students attending high school but not attending higher secondary school. A district-wise analysis of the dropout rate of school education at the secondary level in Mizoram shows that Mamit district scores high compared to other districts, followed by Champhai and Lawngtlai districts.

2.5 Database and Methodology

The UDISE database has been utilized for girls, boys, and overall GER, as well as dropout rates of school education at different levels. The 2011 census database was also used to estimate various age groups in the state's and districts' populations. GERs of primary, upper primary, secondary, and higher secondary schools in different districts were calculated based on enrolment and age population. Descriptive statistics is used to calculate the mean and standard deviation of the school education GER at various levels and the district dropout rates for girls, boys, and overall at the secondary level.

In order to study the reasons for dropout, the 75th round data on Household Social Consumption on Education in India conducted by the NSSO, Government of India, for the state of Mizoram is used. The span of the dataset is July 2017 to June 2018. The survey has been conducted in 29 states and six union territories (UTs) of India. This is the latest dataset available that contains detailed information on education.

The dataset utilized in the study of the implications of student dropout in higher education GER is collected for a period of 19 years, from 2001-02 to 2019-20. This dataset consists of GER for male, female, and total student dropouts in each year, which were neglected by adding the dropout numbers to the enrolment numbers per year. Figures 2.1, 2.2, and 2.3 depict the dataset. Had there been no dropout at different levels of school education, higher education GER would be around 17.7 percent higher than the present GER as of now, which reveals the impact of student dropout on higher education GER.

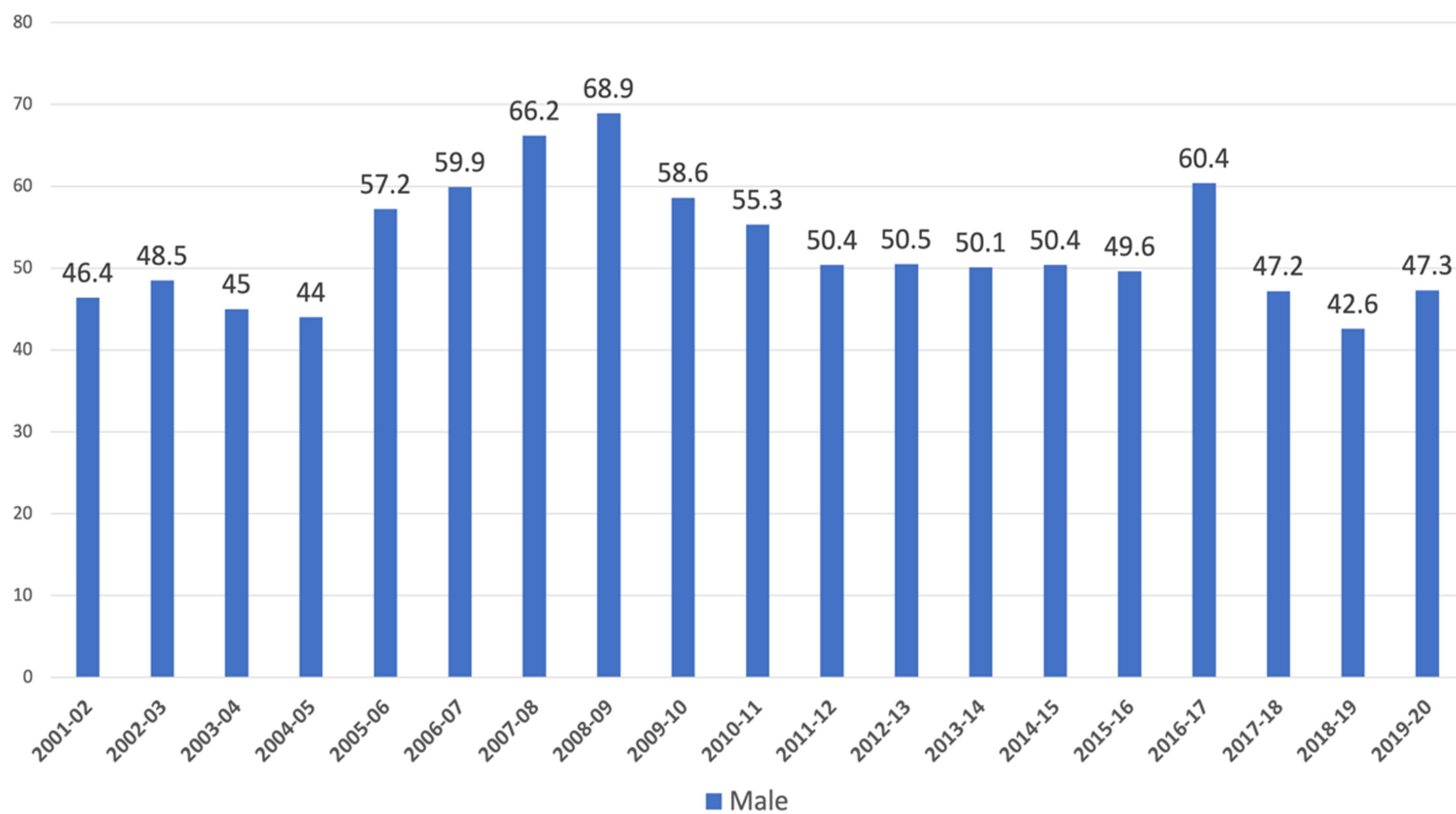

Figure 2.1: Male GER without considering dropout

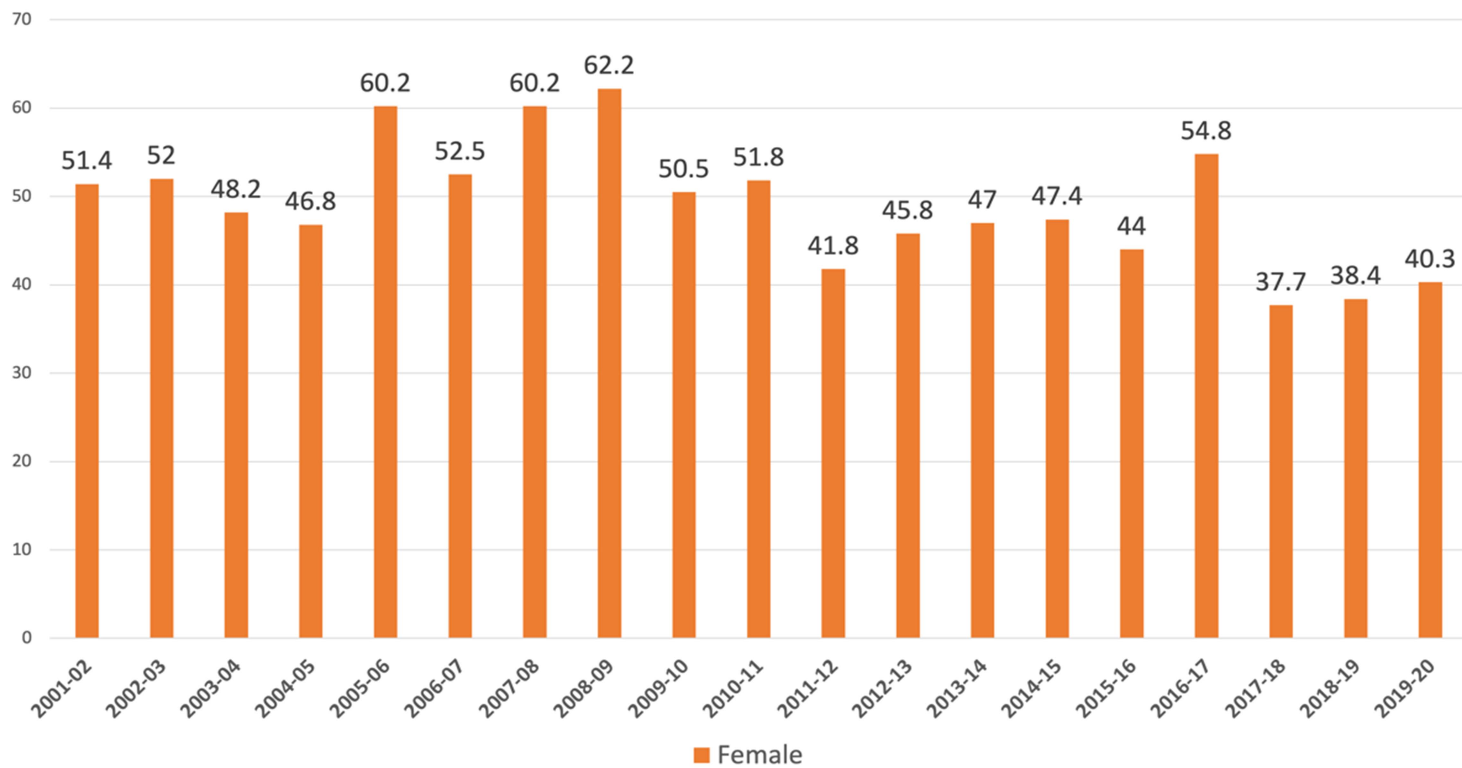

Figure 2.2: Female GER without considering dropout

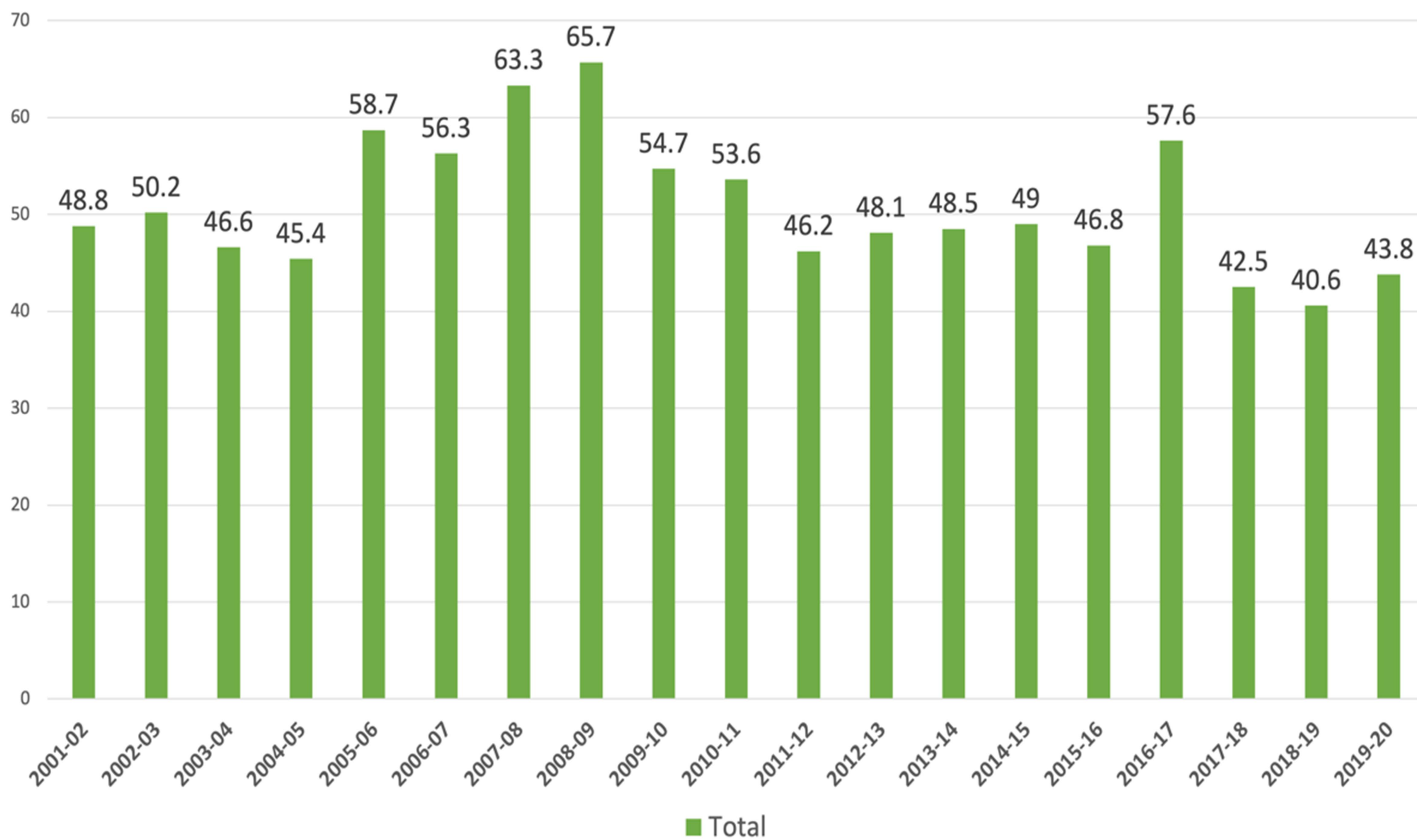

Figure 2.3: Total GER without considering dropout

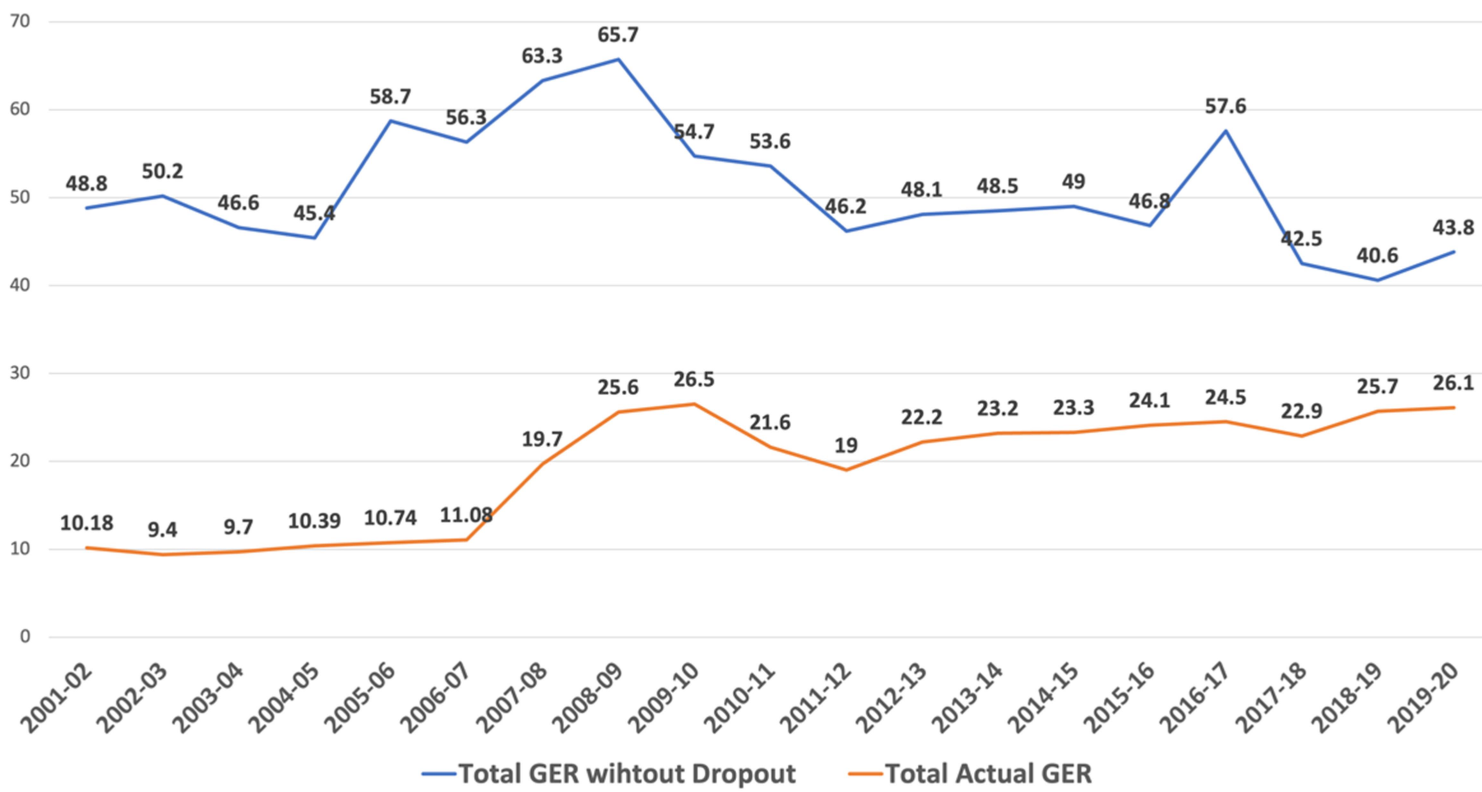

Figure 2.4: GER without Dropout and Actual GER

The techniques implemented in this data analysis and forecasting are described in this section. The state-of-the-art forecasting technique known as ARIMA (Dadhich *et. al*., 2021) is implemented to build a statistical model for total GER. This model is used to forecast future enrolment rates. The overview of the implemented framework can be depicted in Figure 2.5 below:

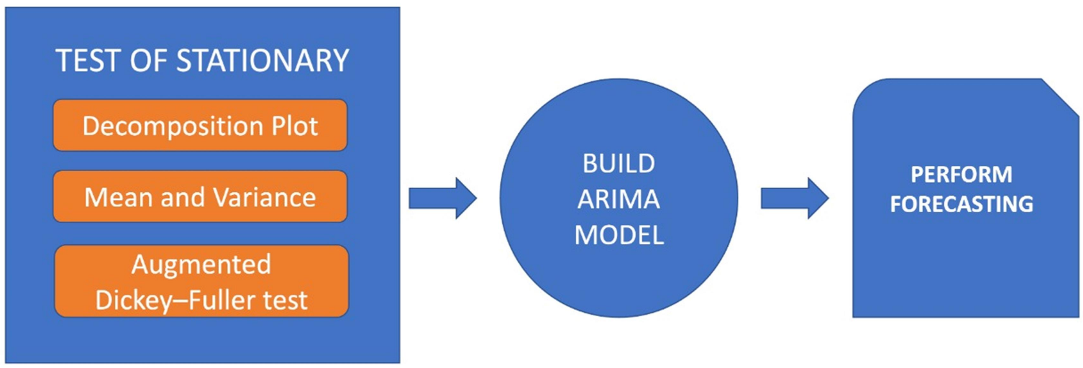

Figure 2.5: Framework overview

2.5.1 Test of Stationary

A key role in time series analysis is played by processes whose properties, or some of them, do not vary over time (Jha *et. al*., 2021). Such a property is illustrated in the following important concept known as "stationary." The common causes of non-stationarity in time series data are the trend and the seasonal components. The way to transform non-stationary data into stationary data is to apply the differencing step. It is possible to apply one or more instances of differencing steps to eliminate the trend component in the data. If the dataset is stationary, it is safe to directly train the model with ARIMA for forecasting in the world of statistics.

2.5.1.1 Decomposition plot

Time series decomposition involves thinking of a series as a combination of level, trend, seasonality, and noise components. These components are:

- Level: It is the main value that goes on average with time.

- Trend: The trend is the value that causes increasing or decreasing patterns in a time series.
- Seasonality: This is a cyclic event that occurs in time series for a short time and causes the increasing or decreasing patterns for a short time in a time series.
- Noise: These are the random variations in the time series.

Decomposition provides a useful abstract model for thinking about time series generally and for better understanding problems during time series analysis and forecasting (Ray *et. al*., 2021). This is one of the tools that can help when encountering a time series problem. This function is similar to a person's vitals being taken when going on a first visit to the doctor. As the vitals might indicate some obvious things in a patient, the decompose plot gives a breakdown of the data and shows if there are any clear trend, seasonality, and the pattern of residuals. Below is the snippet of the code and the output result in Figure 2.6.

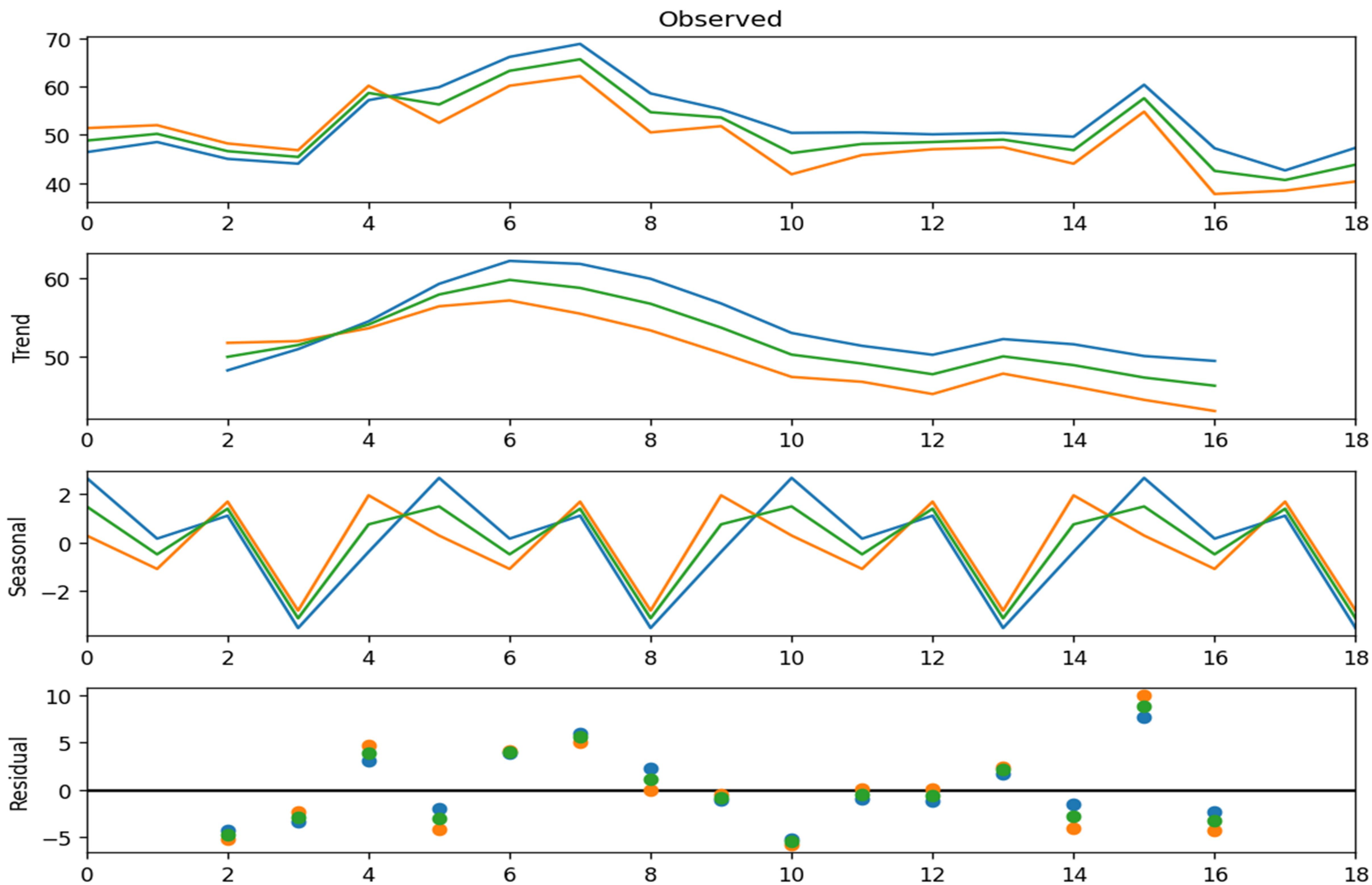

Figure 2.6: Decomposition Plot

2.5.1.2 Mean and Variance

In building statistical model on a Time Series, the series has to be stationary or time invariant, which means that, over different time periods, it should have constant means and constant variance. It means that the data should have constant mean throughout, scattered consistently and should have same frequency throughout. So, if the data mean and variance is varied with time, then the data is non-stationary and it must be made stationary before applying any method. This is necessary because if the data has some regular pattern then there's a high probability that over a different interval, it will have the same behavior and can cause problem in accuracy of model. And also, mathematical computation for stationary data is easier as compared to that of non-stationary data. The dataset in this test is divided into three folds where each fold mean and variance is given in Table 2.10:

Table 2.10: Mean and Variance result

Fold	Mean	Variance
1	51.005556	27.064969
2	55.211111	29.390988
3	52.477778	30.171728

2.5.1.3 Augmented Dickey–Fuller test

ADF test is used to give various values that can help in identifying stationary. The null hypothesis says that a time-series is non-stationary while alternate hypothesis says a time-series is stationary. It comprises of a Test Statistics & some critical values for some confidence levels. If the Test statistics is less than the critical values, the null hypothesis can be rejected and it can be said that the series is stationary (Zhang *et. al.*, 2021). The ADCF test also gives us a p-value. According to the null hypothesis, a lower value of p is better.

2.5.1.3.1 Observations of Augmented Dickey-Fuller test:

Table 2.11: Augmented Dickey-Fuller test

Test Statistic	-4.10
p-value	0.24
#Lags Used	0.00
Number of Observations Used	18.00
Critical Value (1%)	-3.86
Critical Value (5%)	-3.04
Critical Value (10%)	-2.66

2.5.2 ARIMA

ARIMA models are generally denoted as ARIMA (p, d, q) where p is the order of autoregressive model, d is the degree of differencing, and q is the order of moving-average model. ARIMA models use differencing to convert a non-stationary time series into a stationary one, and then predict future values from historical data (de Araújo and da Silva, 2022). These models use “auto” correlations and moving averages over residual errors in the data to forecast future values.

A potential advantage of using ARIMA models is that, as it only requires the prior data of a time series to generalize the forecast, it performs well on short term forecasts. Models are built on non-stationary time series.

2.5.2.1 Prediction

In this section, prediction using the known datasets is performed. 100% of available data is used during the training phase (2001–2019). The training is performed against the known total GER.

The value of p, d, q value used in this implementation is (1, 0, 0).

ARIMA model prediction against the known datasets is given in Figure 2.7.

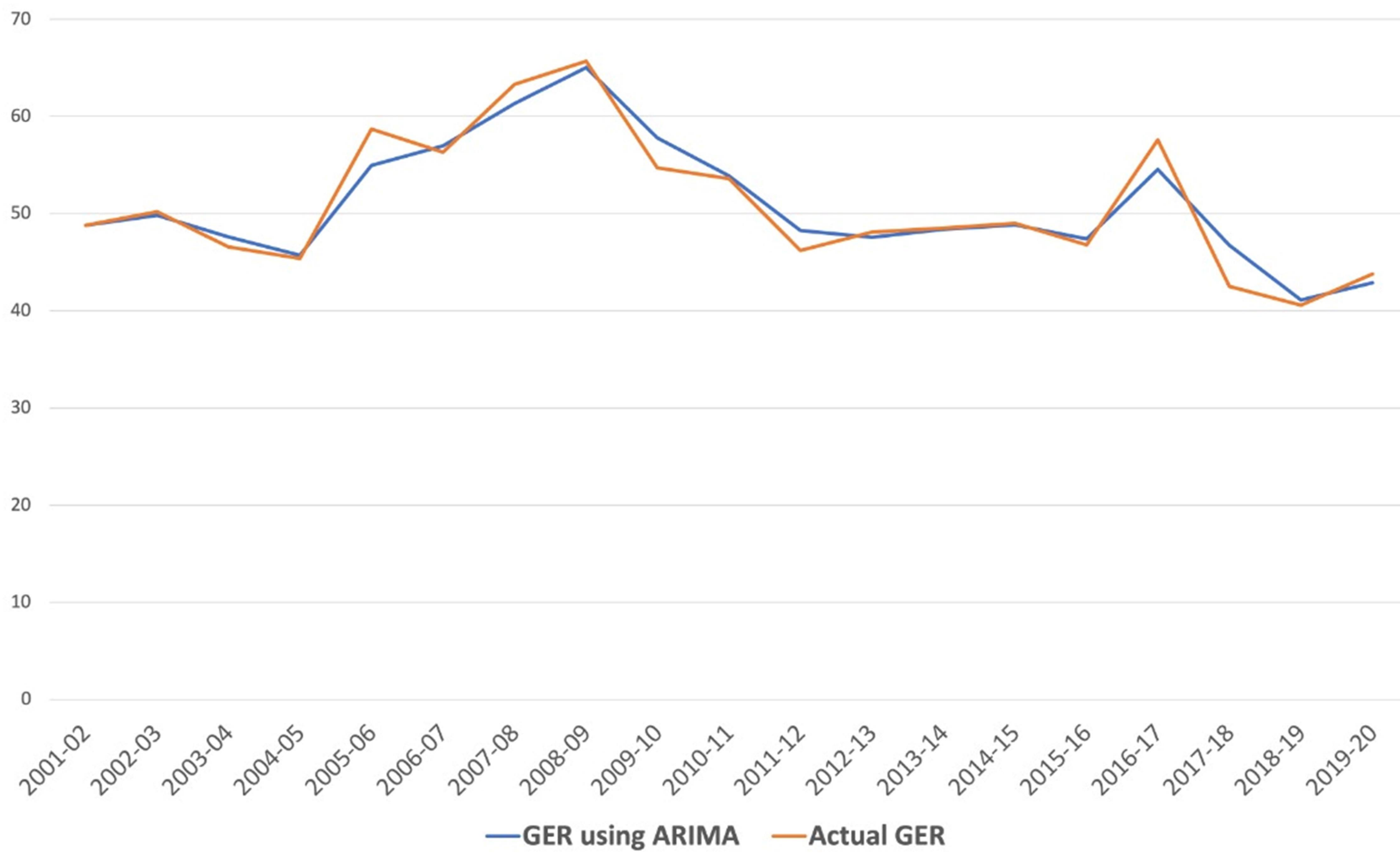

Figure 2.7: Performance of ARIMA

The Mean and RMSE for the above prediction are 50.86 and 1.82 respectively.

2.5.2.2 Forecasting

In this section, forecasting the total GER using the model build in the prediction stage is utilized. The forecasting is performed step wise by taking 10 years model and forecast the next one year. The steps are repeated 5 times so as to obtain the forecasted GER value for the next 6 years, till 2025.

	2020	2021	2022	2023	2024	2025
FORECAST GER	46.99	48.64	49.49	49.93	50.16	50.27

Figure 2.8: GER forecasting without student dropout

2.6 Results and Discussion

In the stationary test section, the decomposition plot Figure 2.6 identified that the datasets of GER time series without dropout are stationary. The graphical plot showing trend, seasonality and residual all favour the characteristic of a stationary dataset. The mean and variance section of Table 2.3 shows a constant mean and variance over male, female and total GER. The dataset is fractioned into three folds, where the entire fold's mean and variance values are relatively close, indicating the stationarity of the GER dataset. Augmented Dickey–Fuller test results in Table 2.11 also indicate that the test statistics value -4.100301 is less than the critical values -3.86, -3.04 and -2.66 at a different percentage. In this case, the null hypothesis can be rejected and it can be concluded that the data is stationary.

As all the above three tests favour the datasets to be stationary, ARIMA is implemented without applying first and second-order differentiation. The implementation is carried out by performing two separate functions: prediction and forecasting. In prediction, all the datasets available (10 years GER) are used for training the ARIMA model. During this training, RMSE is used as an accuracy factor. Value of p, d and q were assigned to minimize the RMSE value. In forecasting, the trained ARIMA model performs GER forecasting for the next ten academic years. The forecasted value is depicted in Figure 2.8.

The GER forecasting without considering student dropout is much higher than the actual GER. However, this high GER rate may take time to achieve. This prediction analysis with minimal student dropout presents the importance of student retention at all levels to achieve higher GER.

2.6.1 Reasons of dropout

As dropout is affecting enrolment, an important concern is the reason for dropout. It may be different from one individual to another. Instead, it is beneficial to conceive of dropout as occurring due to either “push” factors that force a student out of school, or “pull” factors that interfere with a student’s commitment to his or her education (Stearns & Glennie, 2006); (Lan & Lanthier, 2003); (Jordan *et. al*., 1996); (McNeal, 1997) and (Bradley & Renzulli, 2011). After pre-processing the unit level data of NSSO round 75 Schedule 25.2: Household Social Consumption: Education, data is analyzed in order to find out the reason for dropout students in Mizoram on that survey during July 2017 to June 2018. Among the reasons for student dropout, the pull factor, opting out, and other factors all play a signi-ficant role.

Table 2.12: Grouping of the Reasons Provided by 75th Round Schedule of NSSO

Categories	Reasons in 75th Round Schedule of NSSO
Push factor	1. Unable to cope with studies
	2. Unfriendly atmosphere
	3. School is far off
	4. Timings of educational institution not suitable
	5. Language/medium of instruction used unfamiliar
	6. Inadequate number of teachers
	7. Quality of teachers not satisfactory
	8. Route to educational institution not safe
Pull factor	1. Financial constraint
	2. Engaged in domestic activities
	3. Engaged in economic activities
Opted out	1. Not interested to study
	2. Achieved desired level
	3. Preparation for competitive examination
Other	1. Other reasons
	2. No tradition in the community
Female specific	1. Marriage
	2. Female teacher not available
	3. Toilet unavailable

Grouping of reasons for dropping out are explained in the above Table 2.12. It was found that financial constraint, engagement in domestic activities, and engagement in economic activities were the main factors of dropout in Mizoram.

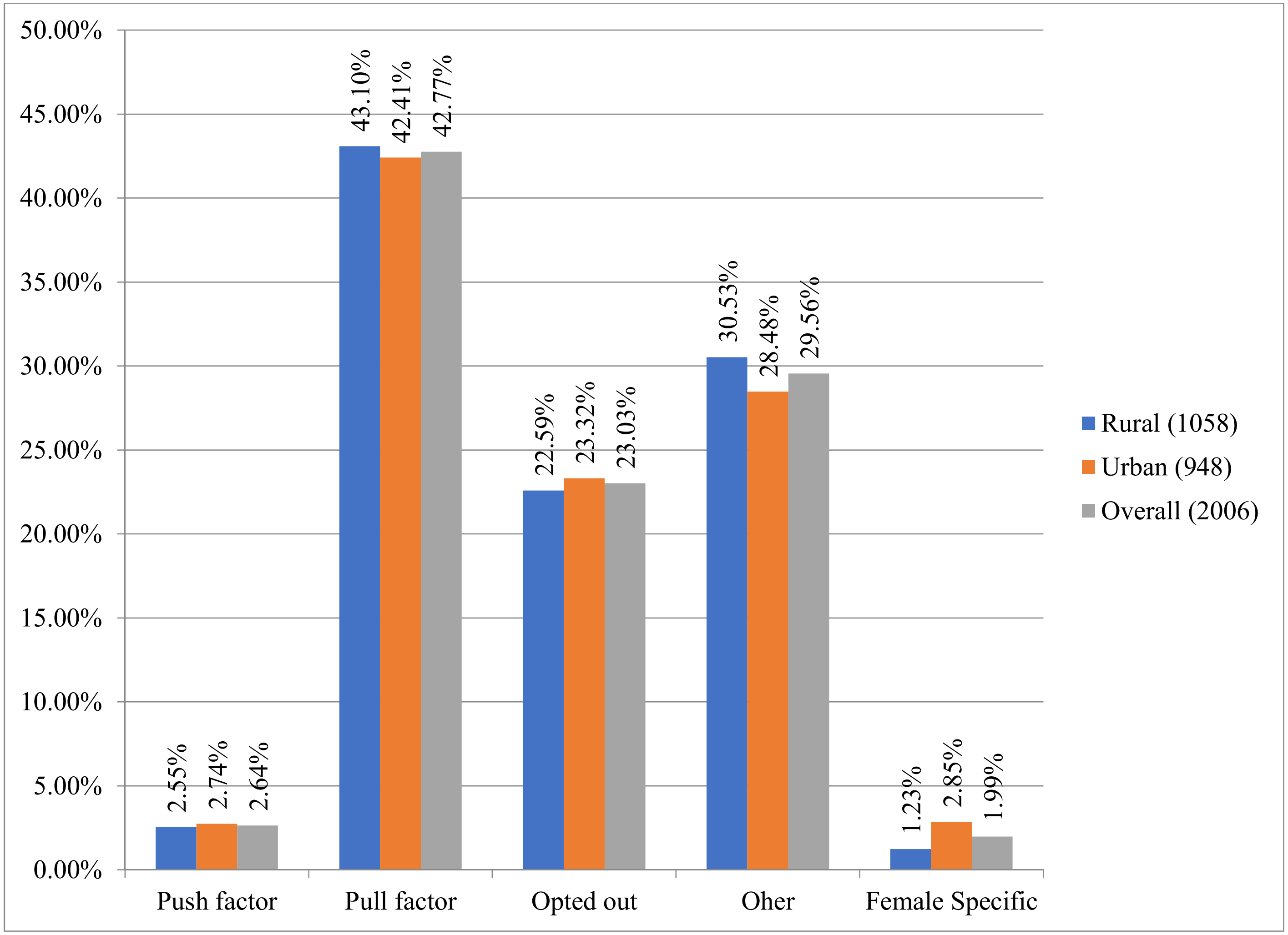

Figure 2.9: Descriptive Statistics of student dropouts in Mizoram

CHAPTER 3

ANALYSIS OF ENROLMENT AND GROSS ENROLMENT RATIO OF HIGHER EDUCATION IN MIZORAM

3.0 Introduction

In India, with the growing size and diversity of the higher education sector, it has become necessary to develop a sound database. It is also required for planning, policy formulation, fulfilling international commitments, research, etc. To address this issue, the Department of Higher Education (DHE) and the Ministry of Human Resource Development (MHRD) decided to conduct the All India Survey on Higher Education (AISHE) on an annual basis beginning in 2010-11, with the following goals: to identify and capture all institutions of higher learning in the country and to collect data on various aspects of higher education from all higher education institutions. The survey covers all the institutions in the country engaged in imparting higher education. To know the real picture of higher education in the country, it is not enough to only know the status of the states and union territories, but also the status of the districts within the states and union territories. This will lead to the identification of educationally backward districts in terms of educational institution, enrolment, gender, social class, and so on.

GER is the total enrolment in higher education, regardless of age, expressed as a percentage of the eligible official population (18–23 years) in the given period. From the definition of GER, a high GER generally indicates a high degree of participation, whether the pupils belong to the official age group or not. GER at each level of education should be based on total enrolment in all types of educational institutions, including public, private, and all other institutions that provide organized educational programs. It has to be noted that GER can exceed 100% due to the inclusion of over- and under-aged pupils or students because of early or late entrants and grade repetition.

UNESCO released the Gender Parity Index (GPI), which is a socioeconomic index designed to measure the relative access to education of males and females. In its simplest form, it is calculated as the quotient of female indicators to male indicators in a given stage of education (primary, secondary, higher, etc.). GPI equal to 1 indicates parity between females and males. In general, a value less than 1 indicates a disparity in favour of boys, and a value greater than 1 indicates a disparity in favour of girls.

A number of literatures have been produced by scholars on enrolment and related aspects. (Adeyami and Akpotu, 2004) critically examined the gender disparity in enrolment in Nigerian universities and discovered a gap between female and male enrolment, with lower female enrolment in all aspects of the universities. They made recommendations, including a sustained enlightenment program, fine-tuning labour laws, and accommodating females under the "educationally disadvantaged" admission policy, to narrow the gender gap in the universities. (Rowan, 2007) studied predictors of delayed college enrolment and the impact of socioeconomic status; he concluded that socioeconomic status is related to timing of college enrolment in the sense that students who enrol immediately or those who delay enrolment have higher socioeconomic status than those who do not enrol. (Sinha and Srivastava, 2008) studied inclusiveness and relative access of social groups to higher education. He found that among the social and religious groups, Muslims fared poorly while high caste Hindus did very well when it came to higher education enrolment, and their pattern of enrolment varied and diversified such that much needs to be done in order to make higher education truly inclusive. (Dubey, 2008) examined the determinants of post-secondary education in India and found that there are large disparities in the enrolment rate between the urban and rural sectors and substantial disparities between the poor and non-poor. (Chuaungo, 2015) studied access to higher education in India and concluded that India still needs to improve access and expand in terms of higher education enrolment and the number of institutions. Data are usually obtained from government and independent publications.

3.1 Institution-wise enrolment in Mizoram

According to the AISHE report 2019-20 (aishe.gov.in), the state has three universities, 35 colleges, and 16 stand-alone institutions. The enrolment of all universities, colleges, and stand-alone institutions in the state is shown in Figures 3.1, 3.2, and 3.3, respectively.

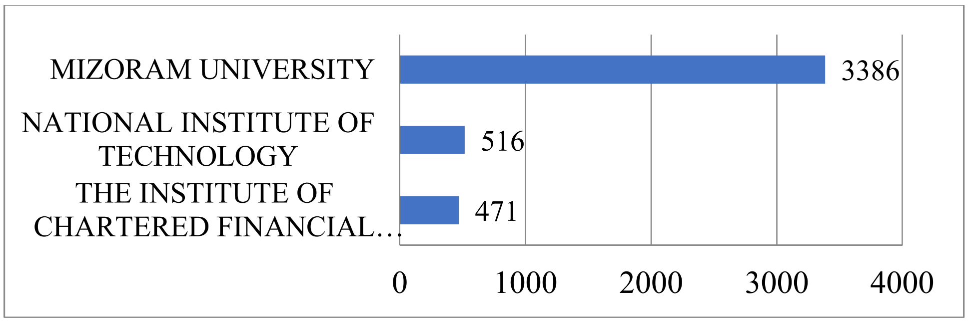

Figure 3.1: Institution-wise enrolment of universities in Mizoram 2019-20

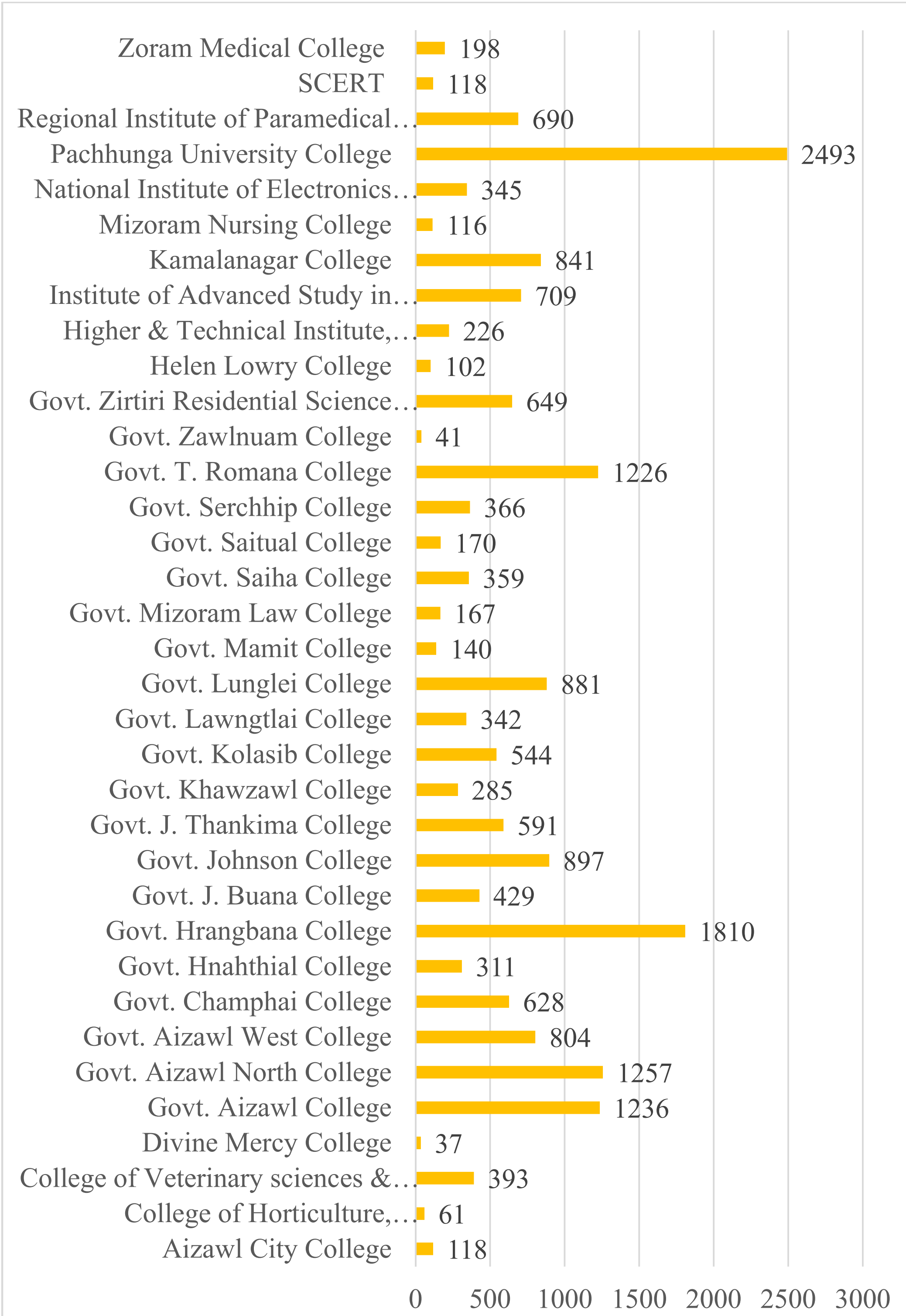

Figure 3.2: Institution-wise enrolment of colleges in Mizoram 2019-20

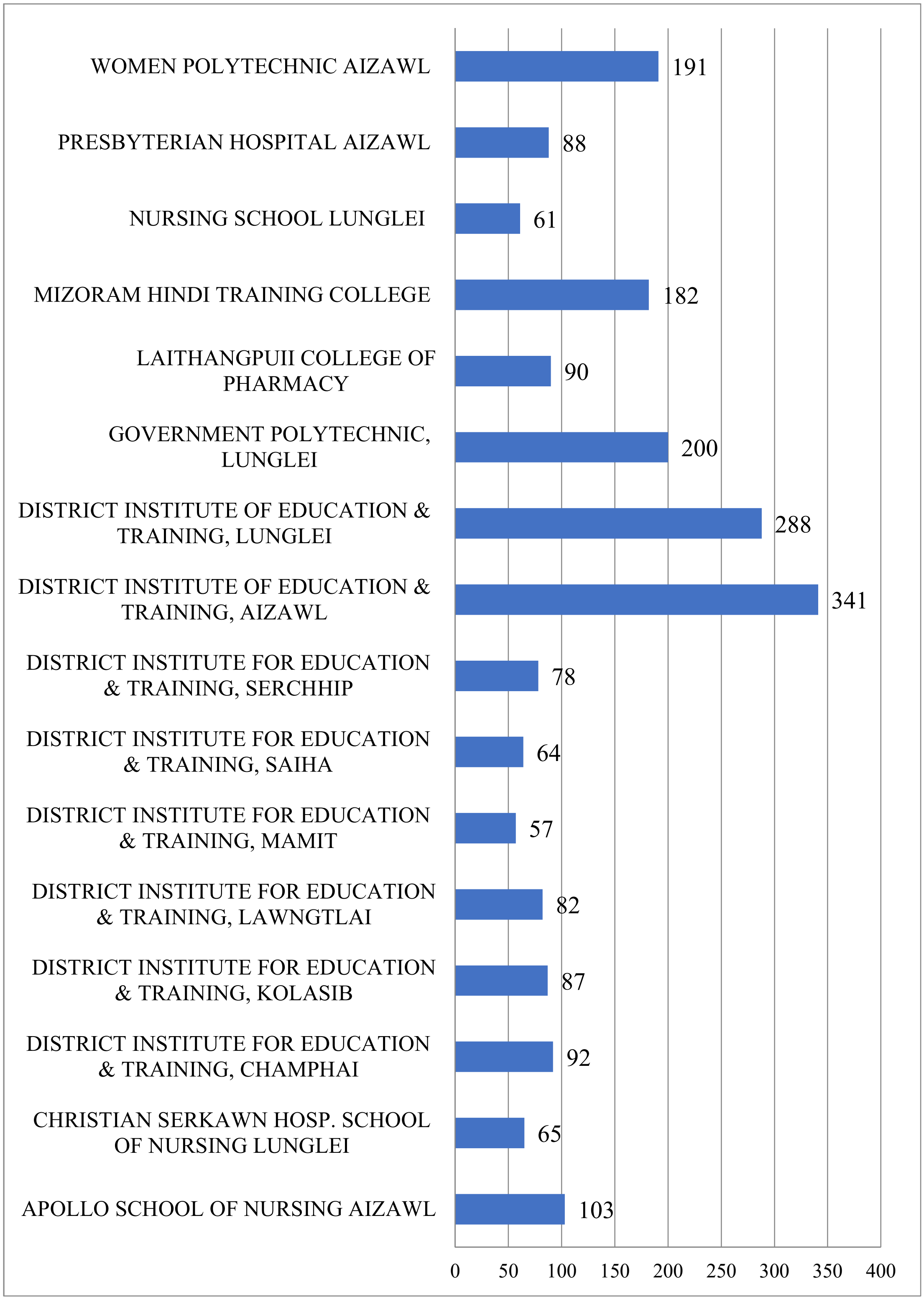

Figure 3.3: Institution-wise enrolment of standalone institutions in Mizoram 2019-20

3.2 Database and Methodology

Several parameters were collected, including the number and type of institutions, teachers, student enrolment, programs, examination results, finance, scholarship and stipend amounts, infrastructure, and so on. Indicators of educational development such as institution density, gross enrolment ratio, pupil-teacher ratio, and gender parity index can also be calculated from the data collected through AISHE.

The following formula is used for calculation:

$GER_t = E_t * 100/P_t$,

Where GER_t= Gross Enrolment Ratio of higher education in the year ‘t’,

E_t = the enrolment in the year ‘t’ and

P_t = the population in age group (18-23 based on last birthday) in the year ‘t’.

Here, we need to know the total enrolment in higher education and the population of the age group (18–23 years) corresponding to higher education. MHRD has been publishing the AISHE report, which contains enrolment data in higher education, on a yearly basis since 2010–11. It also published estimates of population for the years 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, and 2019 in the age group 18–23 years for all states and the Union Territories. Based on the estimated total population in the age group 18–23 years of the state and incorporating the age-wise population from the 2011 census of India, the populations of the districts of Mizoram in the age group 18–23 years are estimated. Thus, the AISHE database has been used to analyze enrolment and calculate GER and GPI for the districts of Mizoram using population estimates as published by the Ministry of Education. The GER for different levels of school education in Mizoram has been taken from the UDISE database (udise.gov.in).

In this study, higher education GER is calculated for different districts for a period of ten years ranging from 2010 to 2019, thereby identifying districts with high and low GER within the state. The Gender Parity Index of Higher Education for all districts is also calculated and analyzed. Enrolment in regular mode and distance mode at different levels of higher education is also analyzed. Furthermore, Pearson correlation was used to investigate the level of relationship between the GERs of higher education and school education.

3.2.1 Pearson correlation

The degree of relationship between GER of higher education and GER of primary school, GER of middle (upper primary) school, GER of high school, and GER of higher secondary school was explored using the Pearson correlation coefficient.

The formula is given below

$$\rho = \frac{\sum_{i=1}^{n}(x_i - \underline{x})(y_i - \underline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \underline{x})^2 \sum_{i=1}^{n}(y_i - \underline{y})^2}}$$

where ρ is the correlation coefficient

x_i= values of the x variables in a sample

$\underline{x}$ = mean of the values of the x-variable

y_i= values of the y-variable in a sample

$\underline{y}$ = mean of the values of the y-variable

3.3 Results and Discussions

3.3.1 Level-wise analysis

Figures 3.4 through 3.12 showed Mizoram's level-by-level enrolment in higher education at various levels from 2010 to 2019:

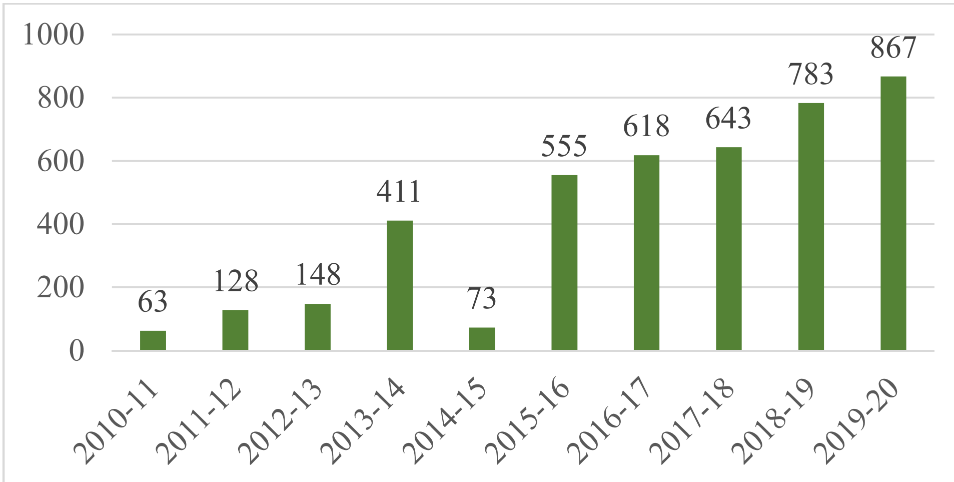

Figure 3.4: Ph.D enrolment in Mizoram

Figure 3.4 revealed that, excluding steep decrease during 2014-15, Ph.D. enrolment progressively increases from 2010 to 2019.

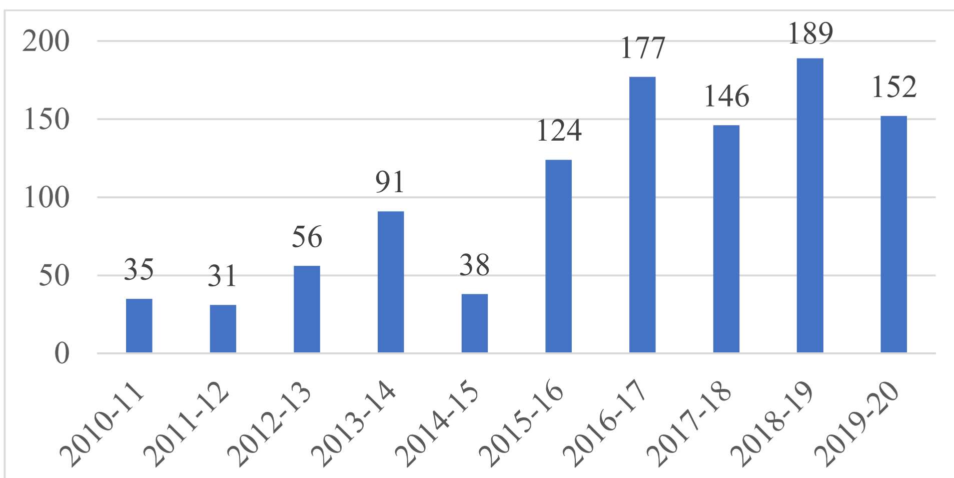

Figure 3.5: M.Phil enrolment in Mizoram

From Figure 3.5, it is observed that M.Phil. enrolment in Mizoram does not follow a regular trend, with the lowest enrolment recorded during 2011–12 and the highest enrolment during the year 2018–19. However, it is seen that M.Phil. enrolment in general increased from 2010 to 2019.

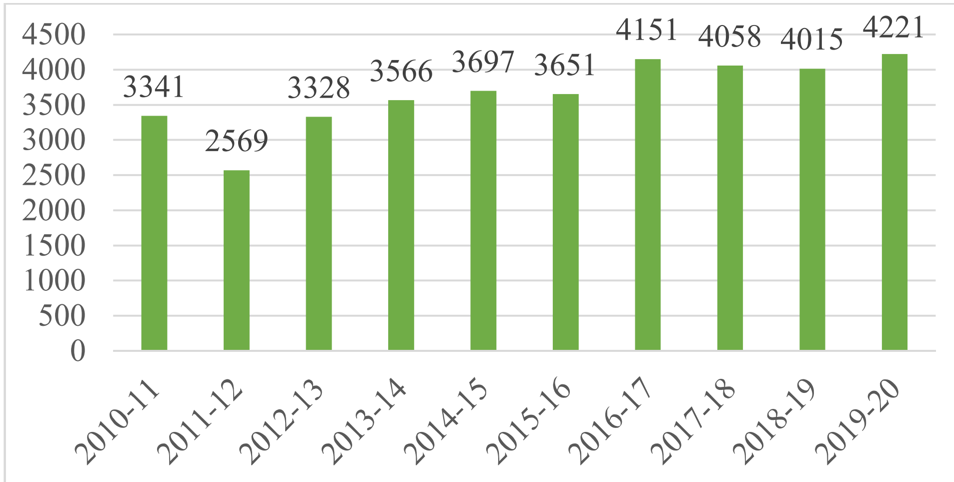

Figure 3.6: Post graduate enrolment in Mizoram

According to Figure 3.6, postgraduate enrolment is seen to be gradually rising during the study period in which the lowest was reported for the academic year 2011–12. However, the enrolment trend from 2016 to 2019 reveals an uneven pattern.

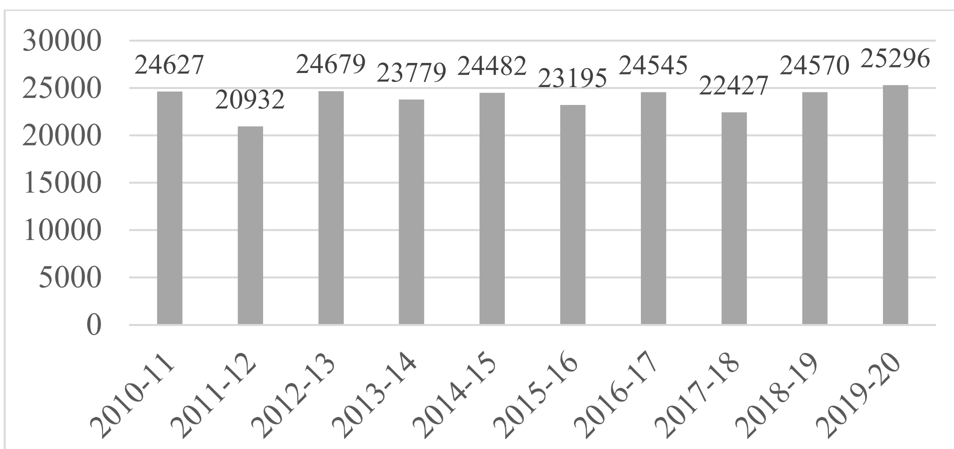

Figure 3.7: Under Graduate Enrolment in Mizoram

Figure 3.7 shows that undergraduate enrolment is increasing five times and decreasing four times from 2010 to 2019. Here also, the enrolment trend is found to be irregular during the study period.

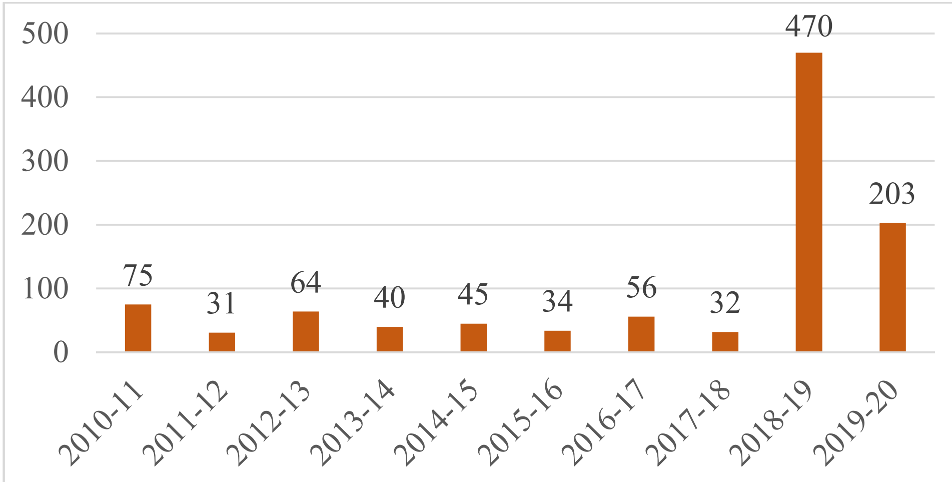

Figure 3.8: Post graduate diploma enrolment in Mizoram

It is seen that there is not much increase in Post Graduate Diploma enrolment from 2010 to 2018. The academic year 2018-19 recorded a steep rise in enrolment but again receded the following year.

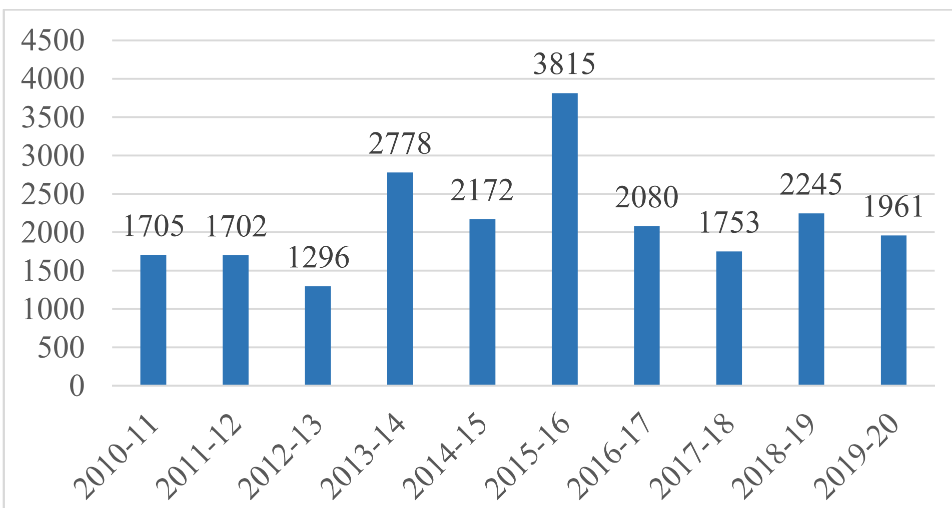

Figure 3.9: Diploma enrolment in Mizoram

Figure 3.9 revealed that there is no regular trend in the enrolment trends for the diploma, which are increasing three times and dropping six times over the study period.

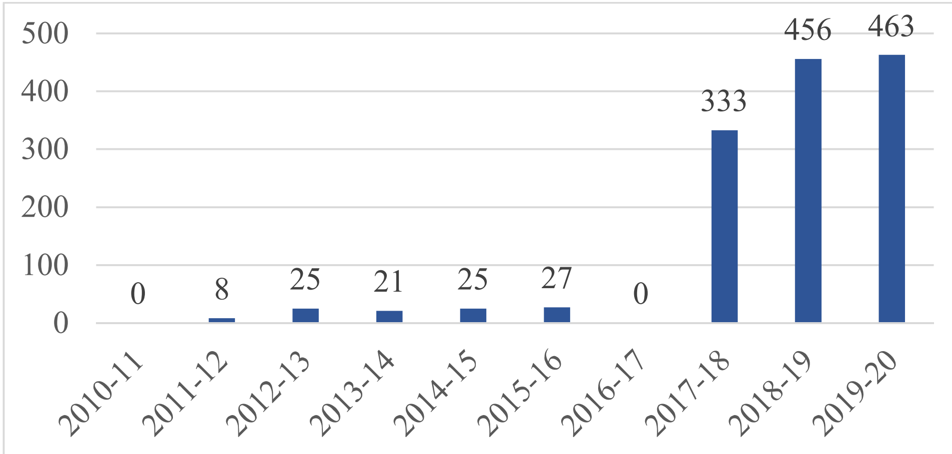

Figure 3.10: Certificate enrolment in Mizoram

Figure 3.10 shows that from 2010 to 2017, not many students enrolled in the Certificate courses. However, from 2018 onwards, these certificate courses received a substantial increase in enrolment.

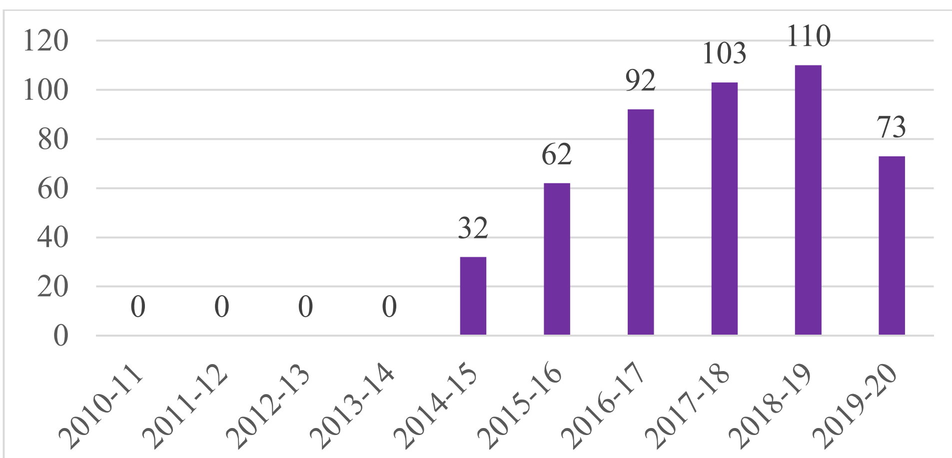

Figure 3.11: Integrated enrolment in Mizoram

From Figure 3.11, it is observed that Mizoram witnessed its first integrated enrolment in the year 2014–15 only, but there is a consistent upward trend until 2018-19.

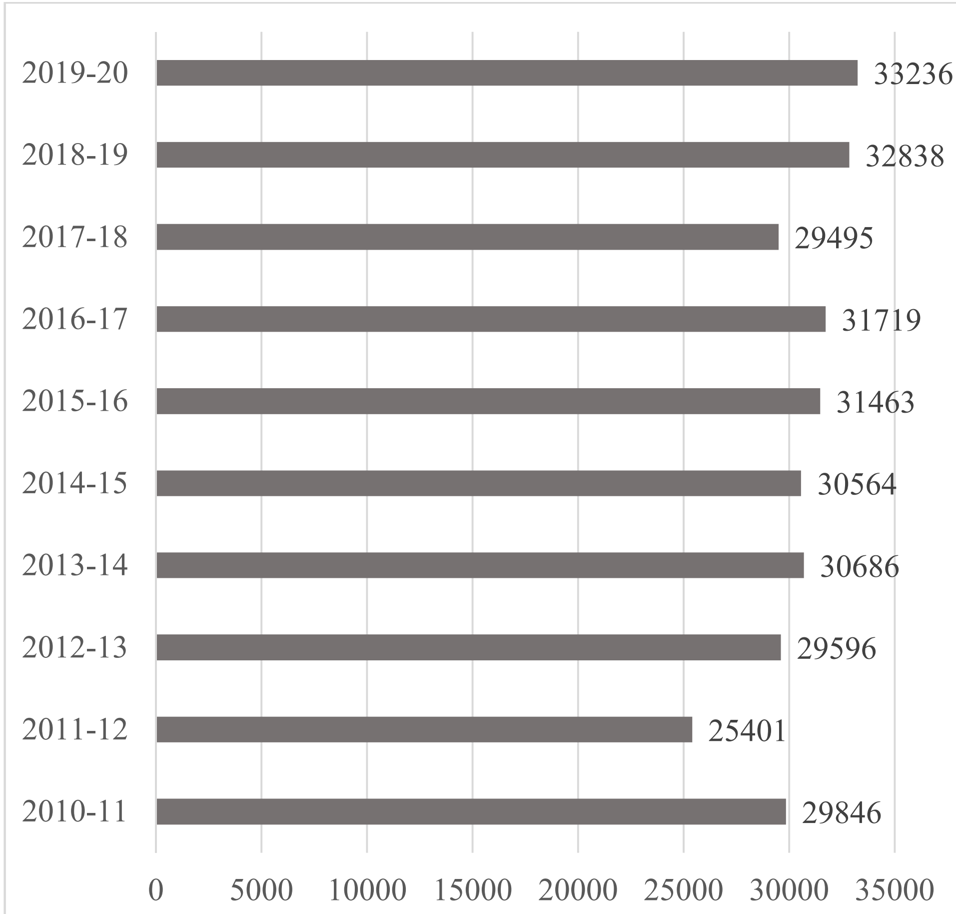

Figure 3.12: Total Enrolment of Mizoram

A careful observation of Figure 3.12 reveals that during the past ten years (2010–2019), higher education enrolment in Mizoram increased from 29846 in 2010 to 33236 in 2019, which is found to be an increase of 11.36 percent.

The analysis of students' enrolment over ten years reveals that undergraduate studies have the highest composition throughout, ranging from 73.32% in 2015–16 to 83.39% in 2012–13, followed by postgraduate studies, which range from 10.11% in 2011–12 to 13.09% in 2016–17. The data has been generated from the AISHE database (aishe.gov.in) and rearranged in a systematic manner. Between 5.37% in 2012–13 and 14.67% in 2015–16, the overall enrolment composition of other studies, including Ph.D., M.Phil., PG Diploma, Diploma, and Integrated Courses, was found to be similar. Table 3.1 gives the total enrolment of districts in Mizoram at various levels during 2019–20.

Table 3.1: Enrolment at various levels of districts in Mizoram (2019-20)

District	Ph.D	M.Phil	PG	UG	PG Dip.	Dip-loma	Certi-ficate	Integ-rated	Total
Aizawl	867	152	4221	19745	203	984	463	73	26708
Champhai	-	-	-	913	-	92	-	-	1005
Kolasib	-	-	-	544	-	87	-	-	631
Lawngtlai	-	-	-	1183	-	82	-	-	1265
Lunglei	-	-	-	1944	-	517	-	-	2461
Mamit	-	-	-	181	-	57	-	-	238
Saiha	-	-	-	359	-	64	-	-	423
Serchhip	-	-	-	427	-	78	-	-	505
Mizoram	867	152	4221	25296	203	1961	463	73	33236

Aizawl district in Mizoram had the highest enrolment in 2019–20 with 26,708 students, followed by Lunglei district with 2,461. Saiha district came in second-to-last with 423 students, while Mamit district came in last with 238.

Aizawl district has the highest enrolment across all levels of study, when study levels are taken into account. Enrolment in undergraduate and diploma programs is present in the other seven districts, but not in Ph.D., M.Phil., postgrad, PG-diploma, certificate, and integrated courses.

3.3.2 District-wise analysis

The district-level GER of Mizoram from 2010 to 2019 was calculated by subtracting the percentage of students enrolled in higher education institutions in that district from the 18-23 year old population of that district, as shown in the table below:

Table 3.2: GER of higher education

	Aizawl			Champhai			Kolasib			Lawngtlai			Lunglei			Mamit			Saiha			Serchhip		
	M	F	T	M	F	T	M	F	T	M	F	T	M	F	T	M	F	T	M	F	T	M	F	T
2019-20	52	50	51	8.9	10	9.5	8.8	8.8	8.8	17	5.5	11	18	17	18	3	3	3.2	7.9	8.4	8.2	7.7	9.8	8.7
2018-19	47	47	47	11	12	11	11	10	10	20	6.6	13	22	20	21	4	4	3.6	9.9	9.2	9.5	9.3	11	10
2017-18	47	41	44	8.8	7.8	8.3	8	8.8	8.4	16	7.5	11	16	15	15	3	2	2.5	10	7.3	8.7	7.9	7.5	7.7
2016-17	49	47	48	12	13	13	10	9.8	10	10	5.4	7.7	14	14	14	3	2	2.6	8.3	7.1	7.7	7.9	8.6	8.3
2015-16	44	42	43	14	15	14	8.2	7.4	7.8	11	5	7.7	23	21	22	7	3	4.9	12	9.8	11	9.7	9	9.4
2014-15	48	48	48	5.2	5.2	5.2	5.2	5.8	5.5	5.4	3.2	4.3	21	19	20	1	1	1.2	6.1	4.7	5.4	5.1	5.2	5.2
2013-14	47	47	47	4.9	4.3	4.6	5	6.5	5.7	4.5	2.8	3.6	26	21	23	1	1	1.3	5.7	5.1	5.4	3.3	3.7	3.5
2012-13	43	43	43	3.9	3.7	3.8	6.4	6.7	6.5	3.4	3.2	3.3	30	26	28	1	1	1.3	2.9	5	4	3.2	3.2	3.2
2011-12	36	35	36	4.2	3.4	3.8	11	8.6	9.6	3.5	2.1	2.8	25	21	23	2	2	1.8	5.9	6.3	6.1	3.1	3.5	3.3
2010-11	41	41	41	9.3	10	9.7	9.8	9.1	9.4	1.3	1.1	1.2	25	19	22	1	1	1.1	12	12	12	3.1	3.1	3.1

Table 3.2 shows that the GERs of Aizawl and Mamit are the highest and lowest, respectively, among Mizoram's eight districts over the years. In terms of the ten-year average GER, Aizawl came in first with 44.8, followed by Lunglei with 21.7, while the other six districts have GERs below ten, and Mamit comes in last with GER 2.4. But, if we check for 2019-20, the latest year under consideration, there are three districts whose GER are higher than ten, namely Aizawl (51.1), Lunglei (17.6), and Lawngtlai (11.1), whereas the GER of the other five districts are below ten, namely Champhai (9.5), Kolasib (8.8), Serchhip (8.7), Saiha (8.2), and Mamit (3.2). The data indicates that district inequality is very high in higher education GER with a maximum of 48.1 and a minimum of 2.6 for the latest year.

Aizawl and Lunglei are first and second in the average male GER of ten years, with 45.6 and 21.9, respectively, whereas the other six districts have GERs below 10, with Mamit at the bottom with GER 2.6.In 2019-20, three districts, namely Aizawl (52.3), Lunglei (18.1), and Lawngtlai (17.0), have GERs above ten. The other five districts, namely Champhai (8.9), Kolasib (8.8), Saiha (7.9), Serchhip (7.7), and Mamit (3.2), have GERs below ten.

Similarly, considering the average female GER of ten years, Aizawl stood first with 44.0, followed by Lunglei with 19.2, while the other six districts are below ten and Mamit is in the bottom with GER of 2.1. But, if we check for 2019-20, the latest year under consideration, there are three districts whose female GER is above ten, namely Aizawl (49.8), Lunglei (17.1), and Champhai (10.2), whereas the female GER of other five districts is below ten, namely Serchhip (9.8), followed by Kolasib (8.8), Saiha (8.4), Lawngtlai (5.5), and Mamit (3.3). The data indicates that district inequality is very high in female GER, with a maximum of 49.8 and a minimum of 3.3 for the latest year.

3.3.3 Gender distribution

Table 3.3 highlights the GPI of Mizoram's higher education from 2010-11 to 2019-20:

Table 3.3: GPI of Mizoram districts

District	2010-11	2011-12	2012-13	2013-14	2014-15	2015-16	2016-17	2017-18	2018-19	2019-20
Aizawl	1	0.97	1	0.99	1	0.95	0.95	0.87	1	0.95
Champhai	1.09	0.81	0.95	0.88	1	1.1	1.06	0.88	1	1.15
Kolasib	0.93	0.81	1.05	1.3	1.12	0.9	0.96	1.1	0.95	1
Lawngtlai	0.85	0.6	0.94	0.62	0.59	0.47	0.53	0.48	**0.32**	**0.33**
Lunglei	0.78	0.83	0.89	0.81	0.92	0.89	0.97	0.9	0.91	0.95
Mamit	0.92	0.94	1.08	1.27	1.18	**0.46**	0.63	0.78	0.99	1.04
Saiha	1.03	1.07	1.72	0.89	0.77	0.81	0.86	0.72	0.94	1.06
Serchhip	1	1.13	1	1.12	1.02	0.93	1.09	0.95	1.19	1.27
Mizoram	0.96	0.93	0.98	0.96	0.98	0.91	0.94	0.85	0.94	0.92

Source: Calculated from AISHE reports

Out of eight districts, six districts of Mizoram except Mamit and Lawngtlai have good GPI during the last ten years. In case of Mamit district and Lawngtlai district, GPI has been very low in some years; 0.32 (Lawngtlai, 2018-19), 0.33 (Lawngtlai, 2019-20) and 0.46 (Mamit, 2015-16) were the three lowest GPI during the period. GPI of the whole state was between 0.91 in 2015-16 to 0.98 both in 2012-13 and 2014-15 which is very good result compared to other states of India.

Gender-wise enrolment of higher education during 2010 to 2019 for the whole state of Mizoram is shown below:

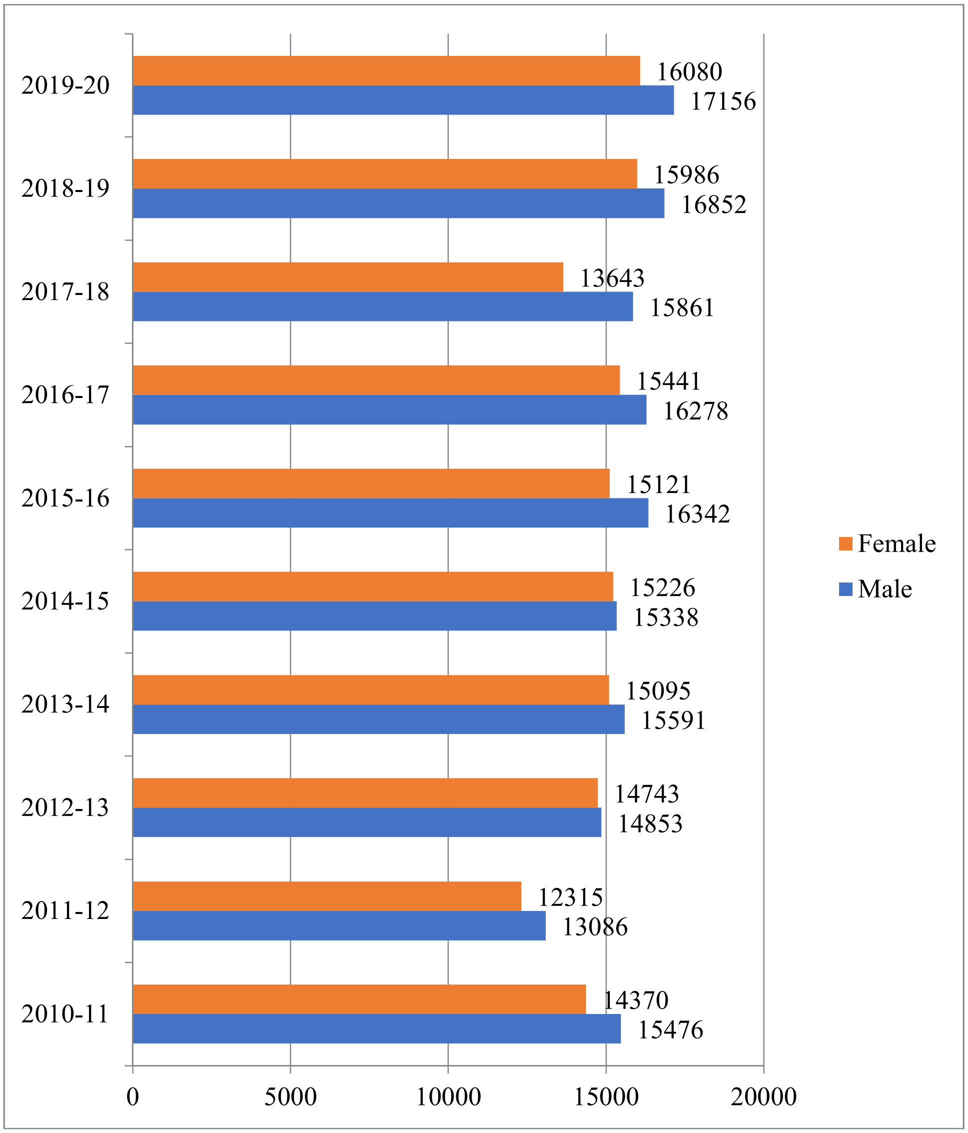

Figure 3.13: Gender-wise enrolment of higher education in Mizoram

If we look into the whole picture during 2010 to 2019, it is fortunate to see that gender differences are negligible in higher education enrolment in the state of Mizoram.

3.3.4 Mode of studies

Table 3.4 showed the enrolment status of distance education in the state of Mizoram during 2010 to 2019:

Table 3.4: Distance mode enrolment of higher education in Mizoram

	Post-Graduate			Under-Graduate			PG Diploma			Certificate			Integrated			Total		Grand Total
	M	F	T	M	F	T	M	F	T	M	F	T	M	F	T	M	F	
2019-20	839	1066	1905	2565	2442	5007	90	113	203	-	-	-	-	-	-	3521	3693	7214
2018-19	702	922	1624	2301	2406	4707	215	255	470	-	-	-	-	-	-	3332	3910	7242
2017-18	1038	669	1707	2085	1200	3285	20	12	32	-	-	-	-	-	-	3147	1897	5044
2016-17	935	1095	2030	1931	1923	3854	43	13	56	-	-	-	-	-	-	3053	3197	6250
2015-16	896	954	1850	1693	1713	3406	23	11	34	-	-	-	-	-	-	3205	3222	6427
2014-15	1083	1068	2151	2074	2211	4285	31	14	45	-	-	-	-	-	-	3521	3695	7216
2013-14	1091	1102	2193	2123	2207	4330	28	12	40	-	-	-	-	-	-	3959	4039	7998
2012-13	1085	1016	2101	2200	2196	4396	58	6	64	-	-	-	-	-	-	3345	3232	6577
2011-12	749	735	1484	1700	1672	3372	23	8	31	-	-	-	-	-	-	2719	2609	5328
2010-11	1274	1075	2349	2463	2175	4638	54	21	75	-	-	-	-	-	-	4045	3452	7497

The number of enrolments in distance mode education is fluctuating between 5,044 in 2017-18 and 7,998 in 2013-14. Most of the students in distance mode are from undergraduate and post-graduate studies.

During 2010-20, the lowest and highest percentage of distance mode enrolment to the total enrolment of the state was 17.10% in 2017-18 and 26.06% in 2013-14. Decline in distance mode enrolment was found for three consecutive years since 2013-14 to 2017-18 but it bounced back in 2018-19. As distance mode of M.Phil and Ph.D were not accepted in India a few years back, this also affected the enrolment figures which can be seen in the following Figure 3.14.

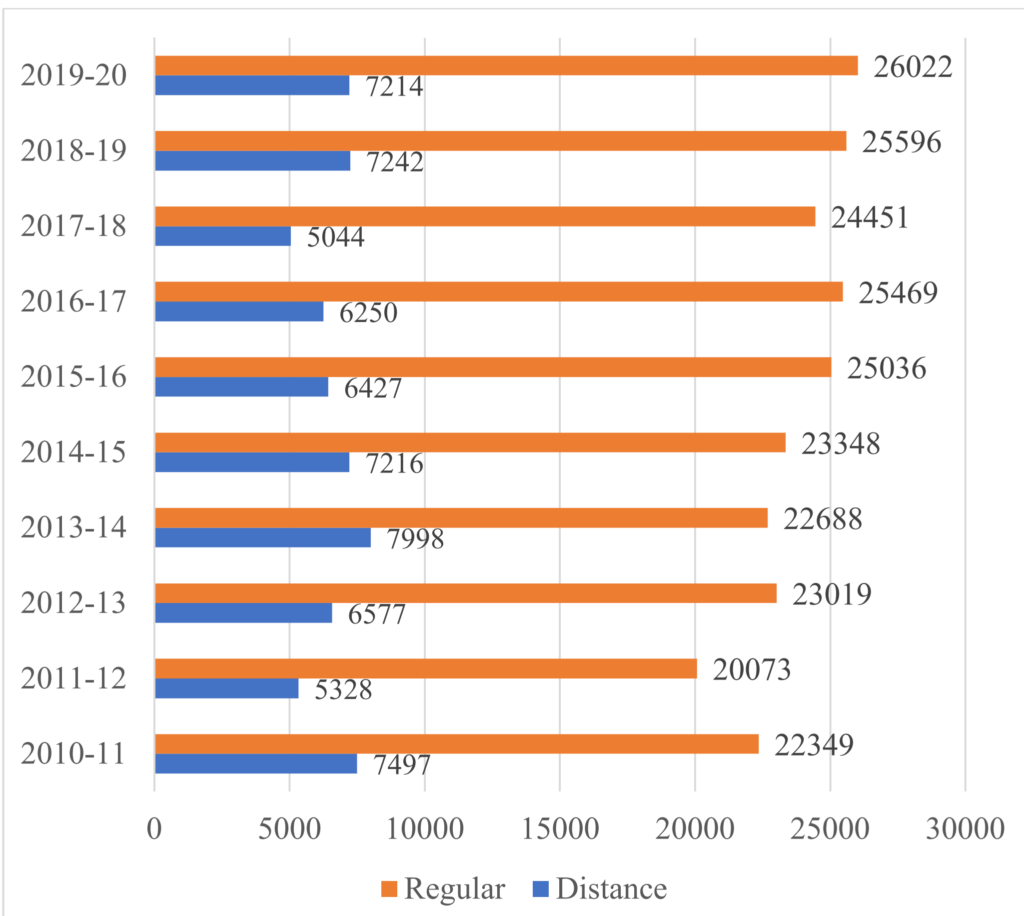

Figure 3.14 – Higher education enrolment of distance mode and regular mode in Mizoram

3.3.5 Correlation analysis

The mainstream educational system in Mizoram is that a student cannot enrol in higher education unless they have completed all four stages of secondary education—primary school, middle school, high school, and higher secondary school. Therefore, it is evident that there is a connection between GER at various educational levels. Using the Pearson correlation coefficient, the levels of association between male, female, and total GER of various educational levels (for N=19) are determined and are shown in Tables 3.5, 3.6, and 3.7.

Table 3.5: Correlation coefficient between Male GERs of different educational levels

GER LEVEL		HE	PS GER	MS	HS	HSS GER
HE GER	Correlation coefficient	1	0.168	0.588	0.784	0.747
PS GER	Correlation coefficient	0.168	1	0.574	-0.086	-0.144
MS GER	Correlation coefficient	0.588	0.574	1	0.53	0.483
HS GER	Correlation coefficient	0.784	-0.086	0.53	1	0.918
HSS GER	Correlation coefficient	0.747	-0.144	0.483	0.918	1

When the male GER of higher education is compared to the male GER of other educational levels, it is found that the correlation between higher education and high school male GER is highest with a value of 0.784, indicating that if high school male GER grows, the male GER of higher education will likewise increase. However, the elementary school male GER is the one with the lowest correlation coefficient, 0.168, showing that among the male GERs, the primary school GER has the least influence on the male GER of higher education.

Table 3.6: Correlation coefficient between Female GERs of different educational levels

GER LEVEL		HE	PS GER	MS	HS	HSS
HE GER	Correlation coefficient	1	-0.005	0.538	0.824	0.852
PS GER	Correlation coefficient	-0.005	1	0.498	-0.178	-0.228
MS GER	Correlation coefficient	0.538	0.498	1	0.508	0.485
HS GER	Correlation coefficient	0.824	-0.178	0.508	1	0.922
HSS GER	Correlation coefficient	0.852	-0.228	0.485	0.922	1

Table 3.6 showed that the correlation between the female GERs of higher education and higher secondary school is highest with a value of 0.852, indicating that if high school female GER grows, the female GER of higher education will also grow. This is true when the female GER of higher education is compared to the female GER of other educational levels. However, the elementary school female GER has the lowest correlation coefficient among the female GERs, at -0.005, indicating that it has the least impact on the female GER of higher education.

Table 3.7: Correlation coefficient between Overall GERs of different educational levels

GER LEVEL		HE	PS GER	MS	HS	HSS
HE GER	Correlation coefficient	1	0.088	0.568	0.814	0.863
PS GER	Correlation coefficient	0.088	1	0.536	-0.13	-0.149
MS GER	Correlation coefficient	0.568	0.536	1	0.522	0.518
HS GER	Correlation coefficient	0.814	-0.13	0.522	1	0.918
HSS GER	Correlation coefficient	0.863	-0.149	0.518	0.918	1

From Table 3.7, it is observed that with a value of 0.863, the correlation between the GERs of higher education and higher secondary school is the strongest, indicating that as higher secondary school GER rises, so will overall GER of higher education. This is observed when comparing the overall GER of higher education to the overall GER of other educational levels. However, the elementary school GER has the lowest correlation coefficient of all the GERs with higher education GER, which, at 0.088, suggests that it has the least influence on the GER of higher education.

3.4 Conclusion

In Mizoram, there has been only a 4.5% increase in GER in higher education from 2010–2011 to 2019–20. The above analysis shows that Aizawl district, which contains the state capital of Mizoram, has the highest GER among the districts of Mizoram. There may be several reasons for this. One of the main factors may be migration to Aizawl from other districts because it is the state capital with better facilities. On the other hand, Mamit district has the lowest GER of 1.1 in 2010–11 and 3.2 in 2019–20, which is very low compared to the overall state GER, which ranged from 19.0 to 26.1 during the same ten-year period. The low GER could also be an indication of student migration from the state to other states in search of quality higher education. Lack of employment opportunity in the skilled sector could be another driving force behind low enrolment in higher education in the state as a whole. So, in order to increase the GER of the whole state, it is required to develop the underdeveloped districts in terms of infrastructure and manpower in the institutions in the districts of Mamit, Serchhip, Saiha, Kolasib, and Champhai.

With the exception of Mamit and Lawngtlai, our results show that the GER of male and female is nearly identical in all districts. This is further corroborated by the increasing GPI for the two districts. This clearly indicates that suitable measures like sensitization through multi-media, seminars, workshops, etc. need to be done to increase female enrolment in these two districts. Further, our study also shows that postgraduate students comprise less than 15% of the total students in higher education, while more than 2/3rd are undergraduate students. Mizoram has only three universities and four colleges that offer postgraduate courses, and many potential students could not get seats in these universities and colleges. It is unfortunate to see that post-graduate and research institutions do not exist in seven districts out of eight districts in the state. Hence, setting up new postgraduate institutions, if possible, is highly recommended, or else strengthening the existing universities and colleges to accommodate more students would also be an important step. Promoting distance learning as a mode of study may also be a good idea, since it has a low financial impact. There are some courses that are not available within the state; thus, setting up an institution depending on the needs of the students is also an important task. To increase access to higher education, there is a need to provide higher education at a lower cost through merit-based scholarship schemes and lower-interest bank loans to underprivileged segments of society. In the meantime, the focus of the administration should not only be on creating new institutes but also on creating employment opportunities so that more students are encouraged to take up higher education.

CHAPTER 4

DATA ANALYSIS AND PREDICTION OF GER USING MACHINE LEARNING AND STATISTICAL METHODS

4.0 Introduction

This section will introduce how machine learning is implemented in the context of Student enrolment analysis and prediction. A common misconception is that predictive analytics and machine learning is the same thing. At its core, predictive analytics encompasses a variety of statistical techniques (including machine learning, predictive modeling and data mining) and uses statistics (both historical and current) to estimate, or 'predict', future outcomes (Hart *et. al*., 2021). These outcomes might be behaviors / approach a system is likely to exhibit or possible changes, for example. Predictive analytics help us to understand possible future occurrences by analyzing the past and present. The overview of this chapter can be depicted in Figure 4.1 as below:

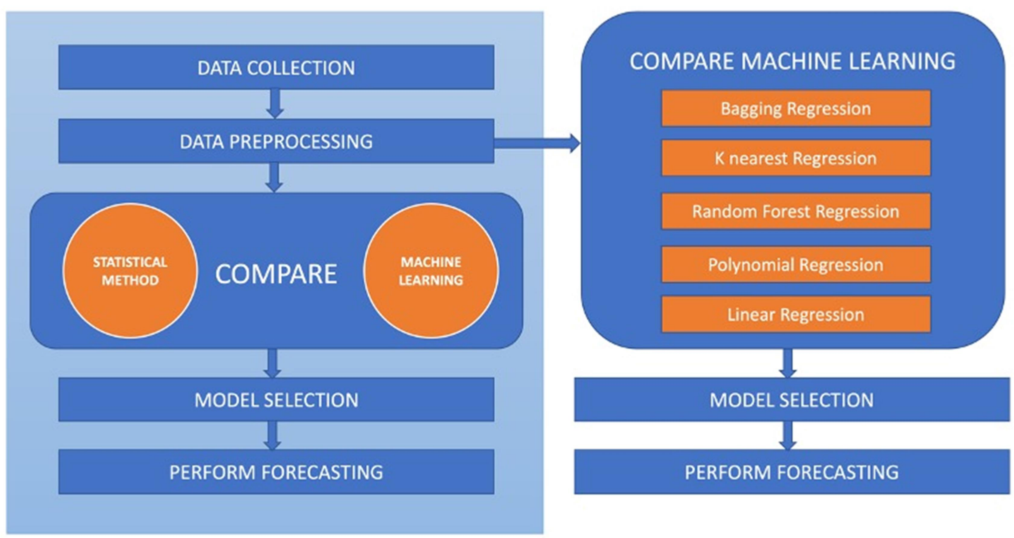

Figure 4.1: Overview of the study

This study is initiated by analyzing and comparing statistical methods (predictive analytics) and machine learning. Comparison is carried out by analyzing the prediction accuracy for the known datasets (enrolment historical data). Based on this comparison accuracy either machine learning or statistical model will be selected. 5-fold validation and accuracy calculated from the confusion matrix are used. Higher accurate model and lower root mean square error (RMSE) are the primary criteria for model selection. The results of this finding selected the machine learning model as it out-performs statistical methods by an average of 0.1322% and 5.6% in both RMSE and Accuracy. Therefore, machine learning is selected for future enrolment prediction.

Future study has been carried out by analyzing different types of machine learning models. In this study Bagging Regression (Ye *et. al*., 2020), K Nearest Neighbour (KNN) Regression (Rahmati *et. al*., 2019), Random Forest Regression (Schonlau & Zou, 2020), Polynomial Regression (Cheng *et. al*., 2018), and Linear Regression (Liu *et. al*., 2019) performance are analyzed. Bagging Regression is found to be the most accurate and again therefore use for prediction of enrolment.

Planned economic development requires data about various aspects of socioeconomic conditions at different levels. Indicators of development are directly or indirectly related to the size and structure of the population (Dziallas & Blind, 2019). It is, therefore, of paramount importance to know various aspects of the size and structure of population at different points in time. Another important requirement of educational planning is enrolment prediction which forms the basis for many of the investment decisions. According to (Rahim *et. al*., 2013), enrolment projection provides information for decision making and budget planning. However, obtaining accuracy is not an easy task, as many factors have impacts on the enrolment numbers. For instance, factors like school fees, politics, and quality of teaching (facilities), strike and security may affect the accuracy and reliability of students' enrolment prediction in Mizoram for higher education. Consequently, students' population and enrolment number should be considered as a chaotic system. In this case, some small change in one of the conditioning factors may bring in a sizeable change in the enrolment figure of students at a given point in time.

Machine learning (ML) is found to be useful in the situations where underlying processes and relationships may display chaotic properties (Lellep *et. al*., 2020). Its applications have been felt in the tasks involving pattern recognition, classification, and time-series forecasting. ML does not require any prior knowledge of the system under consideration and is well suited to model dynamic systems on a real time basis. Unlike statistical and mathematical techniques which depend on influencing factors, the success of ML prediction accuracy depends on parameters adjustment. Therefore, ML is applied to predict the future enrolment of students in Mizoram for higher education. Model was built using a Generalized Feed-forward Neural Network (GFFNN), and its prediction accuracy is compared among the 4 different types of architectures.

4.1. Machine Learning

Machine Learning (ML) refers to a system's ability to acquire, and integrate knowledge through large-scale observations, and to improve, and extend itself by learning new knowledge rather than by being programmed with that knowledge. Machine learning evolved from the study of pattern recognition and explores the notion that algorithms can learn from and make predictions on data. And, as they begin to become more ‘intelligent’, these algorithms can overcome program instructions to make highly accurate, data-driven decisions. The learning process is attained using various algorithms and arithmetic structures to analyze the information. This information is classified by some characteristics called features. ML is used to find a relationship between the features and some output values called labels (Ray, 2019). Mapping of these features and output makes the system to attend a level of intelligence referred to as Artificial intelligence (AI). In order to outperform the traditional approach using AI, training data plays a critical role. However, this dataset is often large and unorganized; so, preprocessing and feature extraction is required. In specific domain such as pattern recognition; deep learning (DL) becomes the state of art where complex training dataset are rendered without bottleneck. This DL enables ML to attend higher level of AI. A board overview of machine learning can be depicted as Figure 4.2 below:

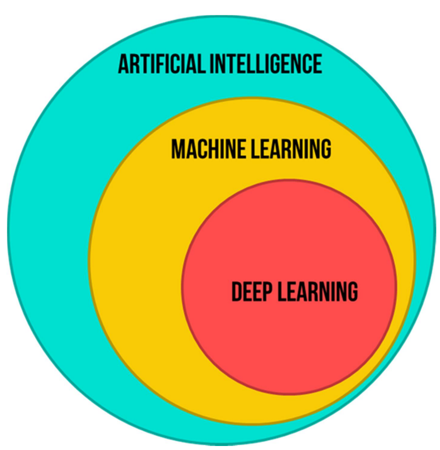

Figure 4.2: Machine learning Venn diagram

Artificial Intelligence is purely a math and scientific exercise but when it becomes computational, it starts to solve human problems.

Machine Learning is a subset of Artificial Intelligence (Jakhar & Kaur, 2020). ML is the study of computer algorithms that improve automatically through experience. ML explores the study and construction of algorithms that can learn from data and make predictions on data. Based on more data, machine learning can change actions and responses which will make it more efficient, adaptable, and scalable.

Deep Learning is a technique for implementing machine learning algorithms. It uses Artificial Neural Networks for training data to achieve highly promising decision making. The neural network performs micro calculations with computational on many layers and can handle tasks like humans.

4.2. Types of Machine Learning

Machine learning is classified into different types, primarily based on how machines learn to solve the given problem and how they solve the problem. An overview of different types of machine learning is depicted as following figure 4.3.

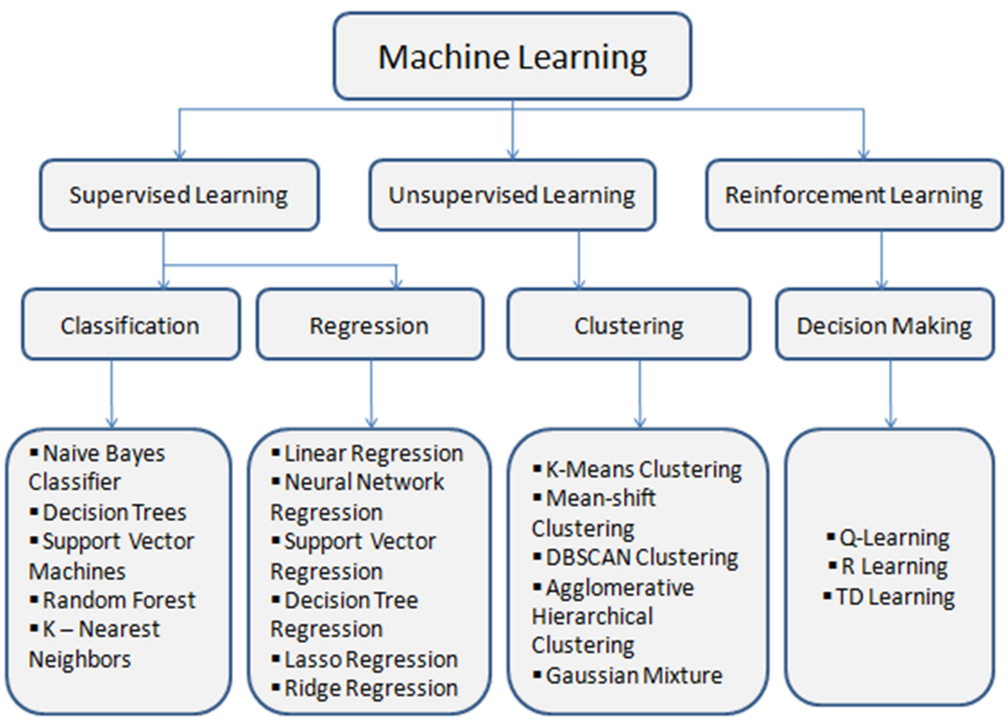

Figure 4.3: Types of machine learning

4.2.1 Supervised Learning

Supervised learning is a technique in which we teach or train the machine using data which is well labeled. The algorithm learns on a labeled dataset, to generate expected predictions for the response to new data. Supervised Learning can be categorised into two types:

4.2.1.1 Classification

In Classification, a computer program is trained on a training dataset, and based on the training it categorizes the data in different class labels. This algorithm is used to predict the discrete values such as male or female, true or false, spam or not spam, etc.

4.2.1.2 Regression

The task of the regression algorithm is to find the mapping function to map input variables (x) to the continuous output variable (y) (Huang *et. al*., 2020). Regression algorithms are used to predict continuous values such as price, salary, age, marks, etc.

4.2.2 Unsupervised Learning

Unsupervised learning involves training by using unlabeled data and allowing the model to act on that information without guidance. Unsupervised learning takes place when the learning system is supposed to detect patterns without any pre-existing labels or specifications. Thus, training data only consists of variables x with the goal of finding structural information of interest, such as groups of elements that share common properties (known as clustering) or data representations that are projected from a high-dimensional space into a lower one (known as dimensionality reduction) (Janiesch *et. al*., 2021). A prominent example of unsupervised learning in electronic markets is applying clustering techniques to group customers or markets into segments for the purpose of a more target-group specific communication. Clustering using unsupervised learning is one of the primitive applications.

4.2.2.1 Clustering

It is the task of dividing the population or data points into a number of groups such that data points in the same groups are more similar to other data points in the same group and dissimilar to the data points in other groups (Sinaga & Yang, 2020). It is basically a collection of objects on the basis of similarity and dissimilarity between them. ie. The data points in the graph below clustered together can be classified into one single group. We can distinguish the clusters as shown in Figure 4.4.

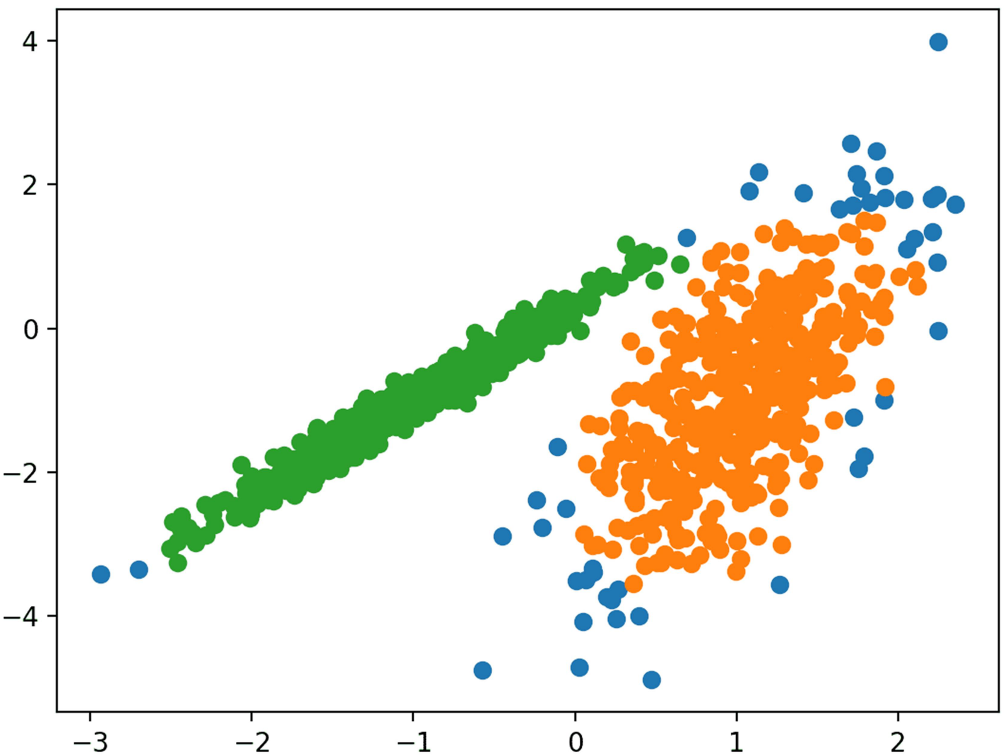

Figure 4.4: Visual representation of clustering

4.2.3 Reinforcement Learning

Reinforcement Learning is a part of Machine learning where an agent is put in an environment and he learns to behave in this environment by performing certain actions and observing the rewards which it gets from those actions (Mazyavkina *et. al*., 2021). It takes appropriate action in order to maximize the positive response in the particular situation. The reinforcement model decides what actions to take in order to perform a given task, which is why it is bound to learn from the experience itself.

4.3 Types of Statistical Methods

Statistical method forecasting implies the use of statistics based on historical data to project what could happen out in the future (Kumar *et. al*., 2022). This can be done on any quantitative data: stock market data, sales, GDP growth, housing sales, etc. In our study we use the time series student enrolment dataset, where student enrolment is generated on a yearly basis.

In the realm of time-series analysis (TSA), autoregressive (AR) processes are modeled using a combination of previous variables, whereas moving average (MA) processes are modeled combining previous forecast errors. They can be merged together resulting in ARMA processes, and generalized to non-stationary processes by differencing the original time series, generating the known as ARIMA (autoregressive integrated moving average) models. Compared to ANNs and other machine learning models, TSA provides a well-established statistical framework that allows drawing conclusions more confidently.

4.3.1 ARIMA

ARIMA stands for Auto-Regressive Integrated Moving Average. It is a model used for statistical analysis of the time-series data. It helps to gain better insights into the data and predict future trends. It works very well with sales data. It is the generalization of the ARMA (autoregressive moving average) model.

- A series is an 'integrated (I)' series if it has to be made stationary using differencing.

A dataset is stationary if it has a constant mean, variance and covariance over time. Differencing is the process of subtracting an observation from the previous time step observation until the data is stationary. There is a mathematical test 'The Augmented Dickey-Fuller test' to determine if the data is stationary.

- The 'autoregressive (AR)' terms are the lags of the stationary series.
- The 'moving average (MA)' terms are the lags of the forecast errors.

Steps to build an ARIMA model:

1. ARIMA models are applied to stationary data only. If the data is not stationary then it is made so by the process of differencing.

2. The autocorrelation and partial autocorrelation patterns are studied to see if there are lags in the data or forecast errors should be added in the forecasting equation.

3. Then the model is fitted and checked for residuals using the ACF and PACF plots. The pattern and the coefficients are checked.

ACF stands for autocorrelation function. The autocorrelation plot is also called a correlogram as it depicts the correlation of the time series where the observations are lagged by k time units.

PACF stands for partial autocorrelation function. It is a conditional correlation between the variables where there is an assumption of values of some other set of variables. If y is the output variable and x1, x2 and x3 are the input variables. Then the partial correlation between y and x3 is determined by taking into consideration the relation of y and x3 with x1 and x2.

The patterns in the ACF and PACF plots can help ascertain the need for AR/MA terms. ARIMA (p, d, q) model can be explained as the non-seasonal ARIMA model is represented using 3 non-negative components. They are:

p = the number of autoregressive terms

It is the autoregressive part of the model and denotes the number of lag observations. The relation between the current observation and the previous period observations are taken into account.

d = the number of non-seasonal differences

It is the integrated part of the model and denotes the number of times the observations are differenced. It takes into account the differencing so as to make the data stationary.

q = the number of moving-average terms

It is the moving average part of the model and denotes the size of the moving average window (order of moving average). It utilizes the relation between observation and residual error from a MA.

When using the ARIMA model, the goal is to figure out which component to use: AR or MA or both, along with how many lags to be used for differencing.

ACF decide the component of the ARIMA model if the autocorrelation plot at the first lag (lag-1) shows positive autocorrelation, then the AR terms are used. If the autocorrelation plot at the first lag shows negative autocorrelation, then the MA terms are used. ACF is the best way to identify a MA model.

PACF relationship with ARIMA model can be described if there is a sharp drop in the plot after the lag 'k' then the AR component is used. If there is a gradual decline, then the MA component is used. PACF is the best way to identify an AR model.

4.4 Methodology

The methodology section can be broadly classified into two sections: the comparison between statistical and machine learning model and the comparison among the 4 architectures of the machine learning model.

4.4.1 Comparative study between machine learning and statistical methods

The overall proposed framework is illustrated in Figure 4.5.

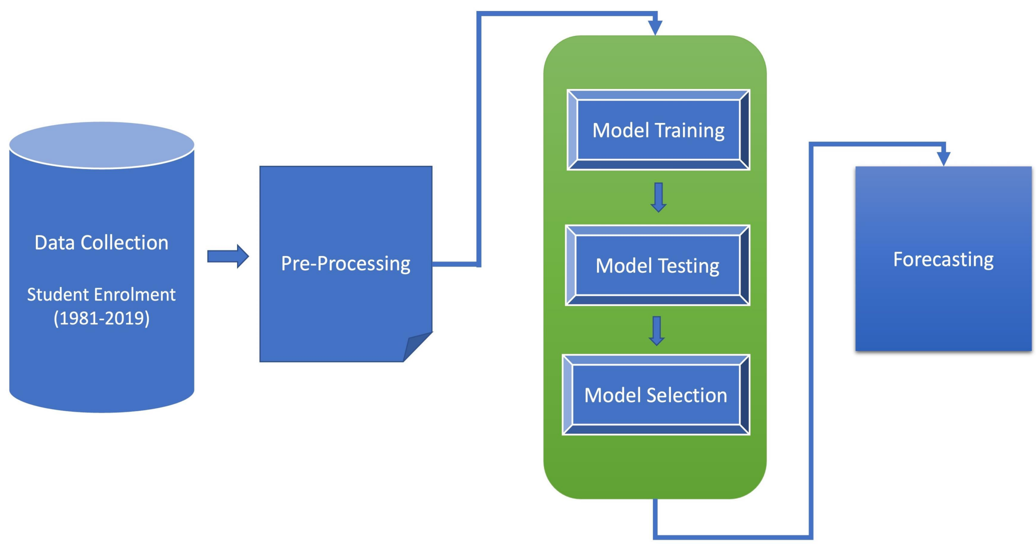

Figure 4.5: Overview of the proposed framework.

The initial stage of the proposed methodology started with data collection, in which the education enrolment data was gathered and analysed. The pre-processing stage deals with data transformation, where year-wise enrolment data is converted to its corresponding GER. The ARIMA and LSTM were trained using the known datasets (GER) in the model training phase. The model testing phase uses the known GER data to test against the prediction accuracy by calculating the accuracy and RMSE. The model selection stage compares and selects the model based on the accuracy results obtained from the previous step to predict the unknown GER (2020–

2050). The forecasting stage performs the actual GER forecast using the selected model.

4.4.1.1 Data Collection

The student enrolment data utilized in this study is acquired from two sources. The enrolment data during the years 1981 to 2009 is collected from Department of Economic and Statistic, Govt. of Mizoram and the year 2010-2019 is from AISHE report issued by the Ministry of Education, Govt. of India. In total 39 years of student enrolment data have been analysed in this study. The datasets is divided into 5 folds, each consisting of 8 years student enrolments from 1st to 4th folds and 7 years student enrolment for the final 5th fold as in Table 4.1. Training and testing have been performed in each fold by taking 70% on each fold enrolment dataset and 30% for testing the models.

4.4.1.2 Data Pre-processing

Data pre-processing is a vital stage in statistical analysis and machine learning, since the arrangement of the data strongly affects the capacity of our model to learn. Intelligent algorithms will not only be enough to generate valuable insights from an inadequate data. Pre-processing stage not only assures the readiness of the data but might also increase the performance of models. In this paper, data transformation is employed in the data pre-processing stage by converting the year wise student enrolment data to its corresponding GER, which is a major indicator for enrolment status. GER represents the number of students enrolled given as a percentage of the population between the age range 18- 23 years. Therefore, GER indicates a ratio of student enrolment based on the eligible population which makes the enrolment and population directly proportional to each other. Figure 4.6 represents data before pre-processing which is the student enrolment data, while Figure 4.7 and Figure 4.8 are the transformed dataset which is GER. Interestingly, while the number of female enrolment is lesser as compared with male enrolment for all the years taken into consideration (1981-2019) as seen in Figure 4.6. But after pre-processing and converting to GER, we can find that for several years such as 1991 to 1996 and 1998 to 2004 the female enrolment is higher than male enrolment as seen in Figure 4.7.

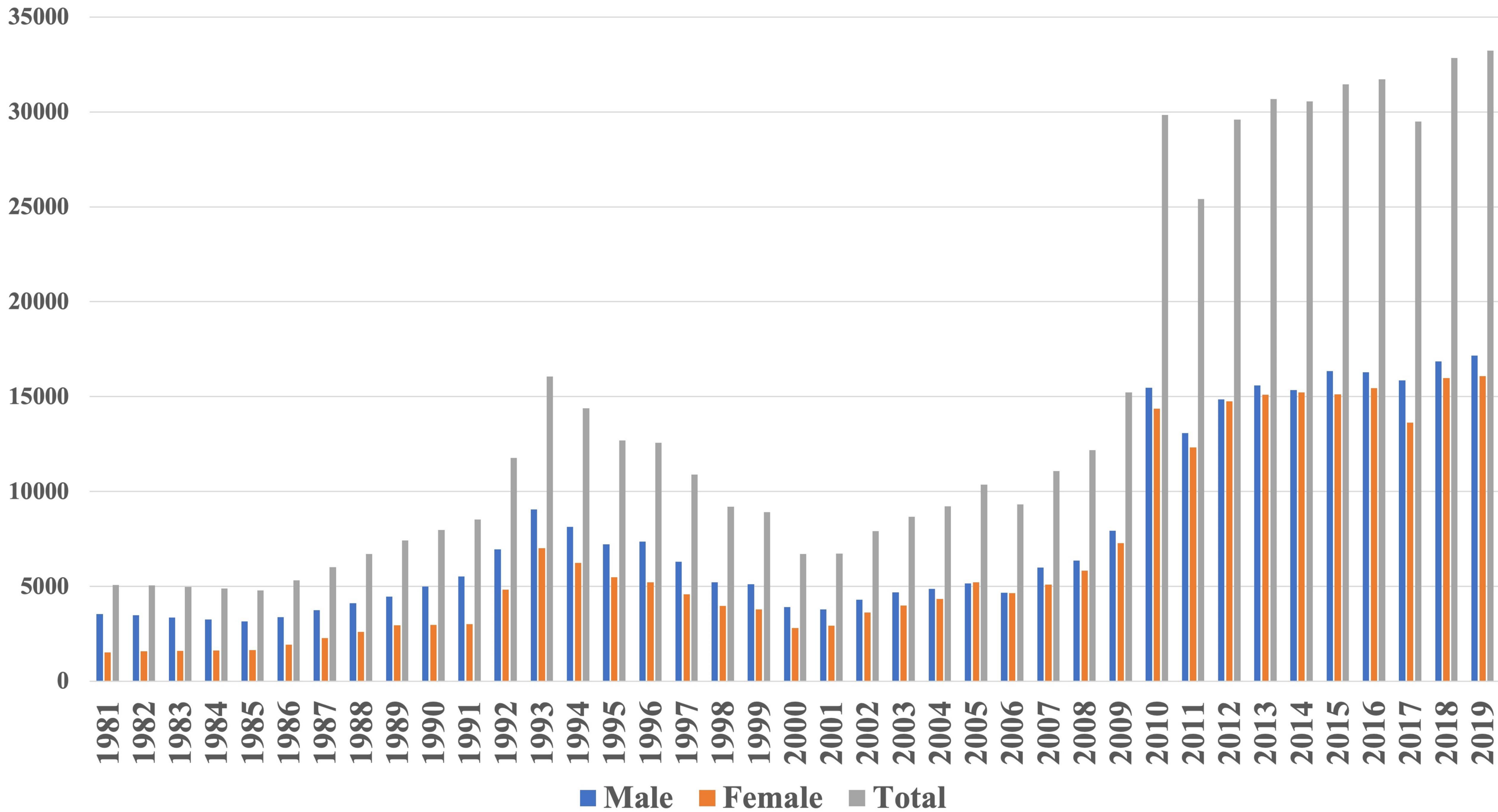

Figure 4.6: Student enrolment (1981-2019)

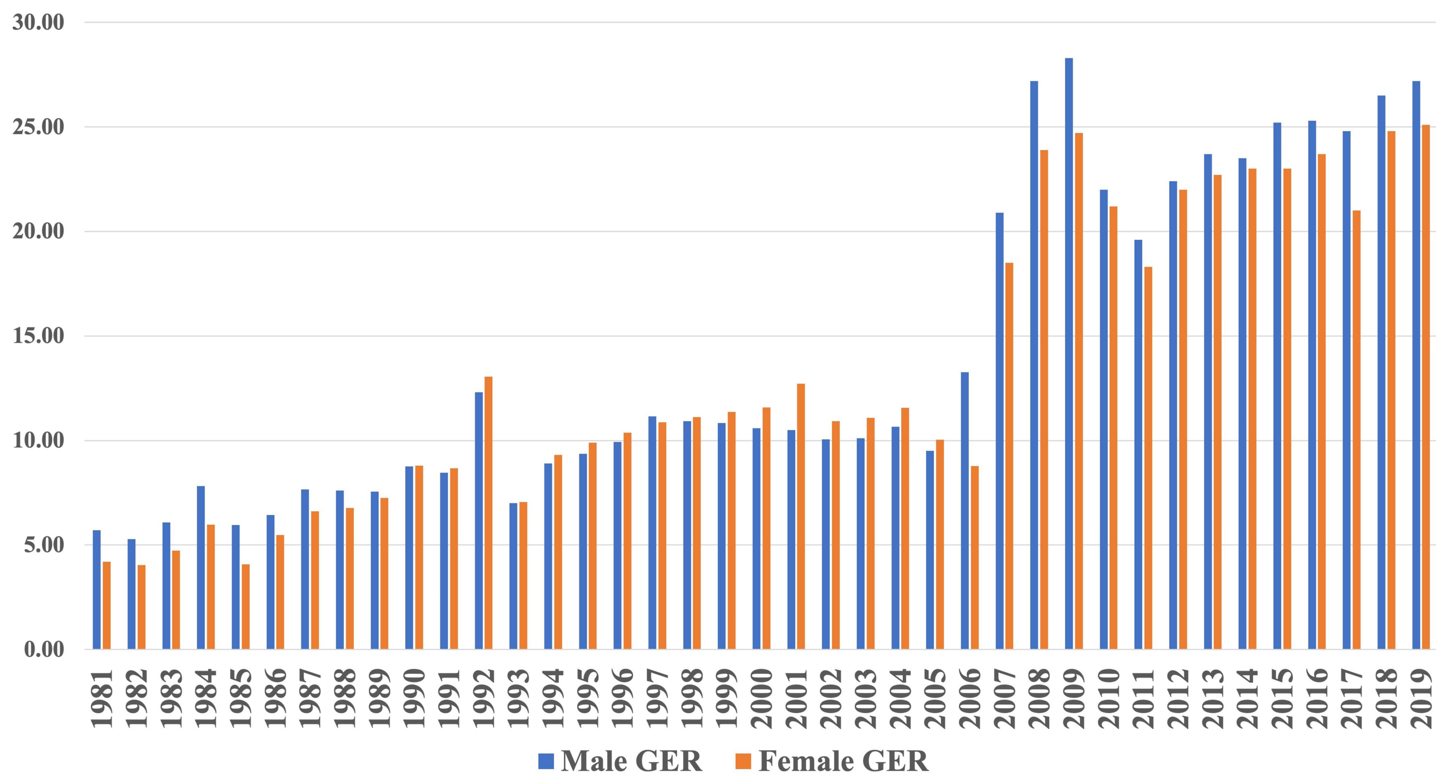

Figure 4.7: Male and Female GER (1981-2019)

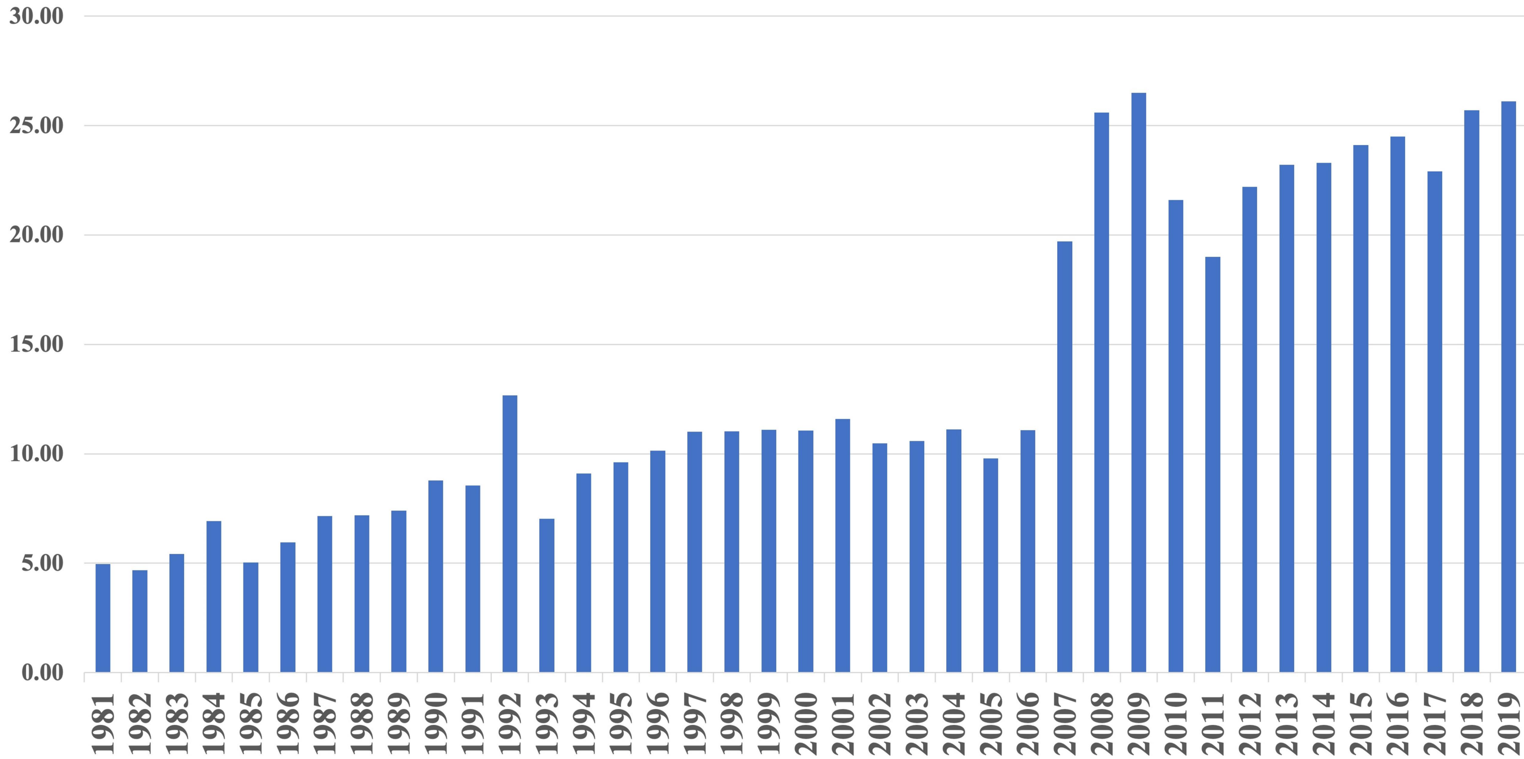

Figure 4.8: Total GER (1981-2019)

4.4.1.3 Model Training and Testing

In our study we compare two state of the art forecasting models known as ARIMA and LSTM. Both these models have gained favour in recent studies both in literature and application. The GER generated by the pre-processing stage is an input to ARIMA, and LSTM for both training and testing. The input datasets are divided into both training and testing as 70% and 30% respectively. Training phase is validated using 5-fold cross validation as this phase is the primary indication for model selection. The model selected in the stage will be used in the testing phase for forecasting GER for the year 2020 to 2035. This training and testing stage is critical as it is the primary determining factor for the how accurate our model can forecast the future GER. Learning weight, bias, epoch for LSTM model are the parameters which vary during the training stage. Likewise the auto regression and moving average functions plays a critical role in ARIMA for determining the model accuracy.

4.4.1.4 Assessment Metrics

In order to identify how the model perform during the testing phase, two popular metrics known as Root mean square error and accuracy were implemented

4.4.1.4.1 Root Mean Square Error (RMSE)

The “loss” is a number that is generated by machine learning. Loss can be defined as a penalty for a poor guess or incorrect prediction. In other words, when the prediction is 100 % accurate then loss value will be 0. Therefore, to reduce the value of loss, two variables named weight and bias of the network are adjusted in every epoch. In addition to loss, researchers most often utilize the RMSE to evaluate the prediction accuracy. RMSE is generally used to calculate the variation between the projected value to the known actual value. The formula is as given in Eq. 1.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \qquad (1)$$

where N is the total number of observation which in our case is the number of years of student enrolment. $y_i$is the actual value of the GER for the known year; while, $\hat{y}_i$ is the predicted value of GER for the unknown year. To improve the prediction accuracy changes over each epoch is calculated using Eq.2 . Often variation in epoch can tune the model to increase the performance while over-fitting is to be taken into consideration. This change is used by the weight and bias of the network during the training phase.

$$Change\ \% = \frac{Predicted\ value - Origianal\ Value}{Original\ Value} X100 \qquad (2)$$

4.4.1.4.2 Accuracy

The accuracy is measured by evaluating the capability of our prediction model by analysing the proportion of properly projected student enrolment. Accuracy is determined from the confusion matrix. This representation can clearly depicted the overall model performance during testing phase. Considering the following Figure 4.9, the accuracy can be determined:

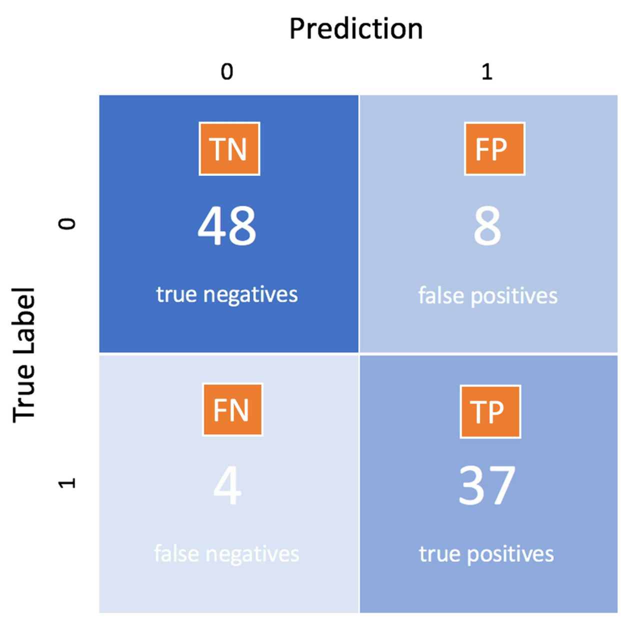

Figure 4.9: Confusion Matrix

Confusion Matrix indicates the accuracy measurement for the model performance. TN and TP are correct prediction while FP and FN are incorrect prediction. The prediction accuracy increases as the value of TN and TP increases.

Hence, accuracy can be calculated using equation (3):

$$\text{Accuracy} = \frac{TN+TP}{TN+TP+FN+TP} \qquad (3)$$

4.4.1.5 Statistical Method (ARIMA)

In statistical method one of the most popular models is known as autoregressive integrated moving average (ARIMA), which is implemented in this study. This model is used extensively to forecast and analyse the time series data. The concept of Autoregressive and Moving Average is combined by integrating it. The auto-correlation and partial auto-correlation are the underlying concepts of ARIMA where AR is based on the concept of partial auto-correlation while MA is based on the concept of auto-correlation (Siami-Namini *et. al*, 2019). The AR part of the ARIMA is based on a dependent connection between the observed data and the delay data (Siami-Namini *et. al*, 2018). The AR of p-order can be defined as the following equation 4.5.

$$X_t = \sum_{j=1}^{p} \emptyset_j X_{t-j} + \varepsilon_t \qquad (4.5)$$

MA: Moving Average: A model which utilizes the dependency between an observation and a residual error of a moving average model applied to delayed data. The moving average process of the q-order or MA (q) is defined as:

$$X_t = \varepsilon_t - \sum_{j=1}^{p} \theta_j \varepsilon_{t-j} \qquad (4.6)$$

Generally, the ARIMA model is represented as ARIMA (p,d,q), where p characterizes the order of the autoregressive process, d defines the order of the stationary data and q indicates the order of the moving average process. The ARIMA model can be mathematically expressed by:

$$(1-B)^d y_t = \frac{\theta(B)}{\emptyset(B)} \varepsilon_t \qquad (4.7)$$

The terms AR and MA can be described as equation 8.

$$\emptyset(B) = 1 - \emptyset_1 B^1 - \emptyset_2 B^2 - \cdots - \emptyset_p B^p$$

$$\theta(B) = 1 - \theta_1 B^1 - \theta_2 B^2 - \cdots - \theta_p B^p \qquad (4.8)$$

4.4.1.5.1 Listing 1: ARIMA implementation

```
Inputs      : GER (Time series data)
Outputs     : RMSE for the forecasting data
1.    nsize ← length (GER) * 0.60
```

```
2.    training ← GER [0...nsize]
3.    testing ← GER [size... length (nsize)]
4.    historyData ← training
5.    predictionsData ← empty
6.    for each p in the range (lenght (historyData)) do
7.        model ← ARIMA (historyData, order=(4, 1, 0))
8.        modelFit ← model.fitting()
9.        forecasting ← modelFit.forecast(predictionsData)
10.       predictions.append(forecasting)
13.   end for
14.   MSE =Find_MeanSquareError(historyData, forecasting)
15.   RMSE = Find_sqrt (MSE)
16.   Return RMSE
17.   Return Accuracy
```

4.4.1.6 Machine learning (LSTM model)

Learning big dataset with large variety of dependent variable was difficulty to be model accurately using the predecessor network known as RNN. Therefore an extension version of RNN was created which can solve the vanishing gradient problems. Large dataset are often accompanied with Larger dependency, to manage this dependency LSTM was introduce. LSTM are built with memory type structure which can remember the previous state and take decision based on the information from past dataset (Bousnguar *et. al*., 2022). In other words, this extension version of RNN has the ability to learn sequential data by retaining information of all the relevant previous stages. Therefore, the memory introduce in LSTM makes it superior as compare with the predecessor RNN. In LSTM the choice of whether to retain or delete the information about the previous stages are controlled by a cell known as “gates”. This gates perform by analysing the weight value assigned during the training phase of dataset. If the associated weighted value is lower than threshold the previous phase information are deleted, if not then the information is retained.

Generally, LSTM architecture is build using three gates known as: Input, Forget, and Output gates. The forget gate is responsible preserving or deleting the information, while the input gate determine which information will be inserted into the LSTM memory. Finally, the output gate check in each epoch, weather the cell value makes any significant contribution to the model output. The overall gates can be depicted as Figure 4.10.

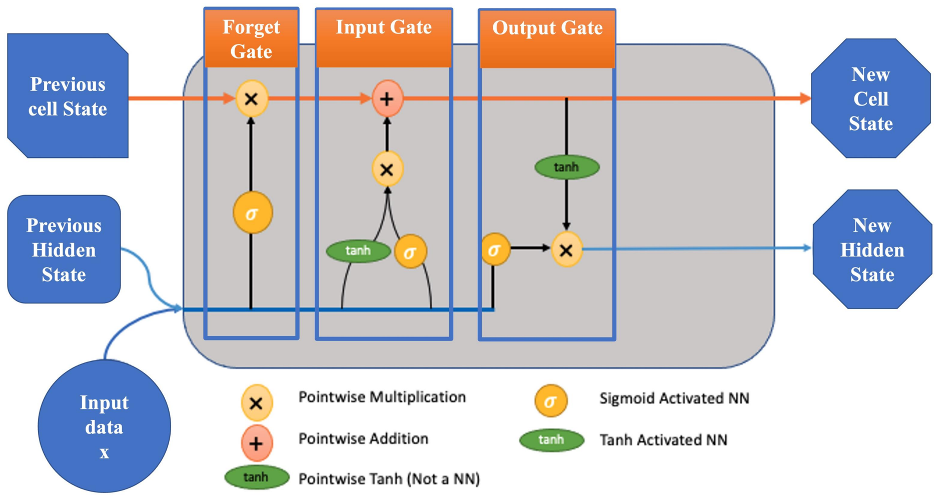

Figure 4.10: LSTM Gates representation.

The LSTM gate representation shows a graphical diagram of how the input data is processed. The unique feature of this model is how the previous input state contributes to the current input state in each gate. Input data is combined with the previous cell state with functions such as pointwise multiplication, pointwise addition, pointwise Tanh, sigmoid activation, and Tanh activation. Each of these functions is graphically represented in different colours. The line chart also gives the data flow along the LSTM path. This line chart is again classified into forget, input, and output gates. Details of these gates are as follows:

(a) Forget Gate: The first steps in LSTM known as the forget gates it decides which state of cell is important in considering both the hidden state and the input data. To make this gate functional both the new input x_t and previous input cell state h_{t-1} are trained using sigmoid function. The output of this always between the interval of [0 and 1]. When the value is closer to 0, the information is consider not relevant and closer to 1 means retaining more information. In other words when the output value f_t is 0 the network tends to forget and when the output is 1, the information is retained. This output of Forget gate can be represented as following equation 4.9:

$$f_t = \sigma(Wf_h[h_{t-1}], Wf_x[x_t], b_f) \qquad (4.9)$$

where b_f is a constant with fixed value throughout the epoch and is also often known as bias value.

(b) Input Gate: This gate is responsible for adding new information to the cell state by considering the value of the given previous hidden cell and the new data which is input. The input gate is a combination of two function known as sigmoid Activation and $tanh$Activation. The sigmoid will make a decision on which value have to be change in cell. As sigmoid output a value of [0 -1], where 1 means allow to change or update all data whereas 0 means not allow to change or update. the "$tanh$" layer on the other hand output a value between [-1,1], where it represent a candidate value from the cell state which will be added to a new memory cell of LSTM. The input gate output can be mathematically represented as following equation (4.10) and (4.11) as follows:

$$i_t = \sigma(Wi_h[h_{t-1}], Wi_x[x_t], b_i) \qquad (4.10)$$

$$\tilde{c}_t = tanh(Wc_h[h_{t-1}], Wc_x[x_t], b_c) \qquad (4.11)$$

Output of Equation (10) i_t pass through a sigmoid function and tell us the old value is to be updated or not. While equation (4.11) output $\tilde{c}_t$ gives a list of vector to the newly candidate list that will be inserted into a new memory cell. These two gate (forget and input) work in synchronized manner and update the new memory cell state accordingly.

(c) Output Gate: The output gate also known as the final gate is responsible for deciding which will be the new hidden state. The input to this gate is a combination of previous hidden state, updated cell state and the newly input data.

A non-linear function known as **tanh** function is applied which generate a value between -1 and 1. This output is then multiplied to the output of the sigmoid layer. Equation (4.12) and (4.13) give a mathematical representation as follows:

$$o_t = \sigma(I[h_{t-1}], W0_x[x_t], b_0) \qquad (4.12)$$

$$h_t = o_t * \tanh(ct) \qquad (4.13)$$

where output value LSTM is given by o_t , and the representation value between -1 and 1 is given by h_t.

The implementation of LSTM algorithm is described in Listing 2. Initially data were separated into 30% and 70% for both training and testing separately. A function call fit LSTM is used for training the algorithm with an input of training dataset, number of epoch and number of neurons. Line 8 and 9 set the model parameters by using RMS Prop as optimizer, loss is computed in MSE. Line 10 to 13 accomplishes training the actual LSTM model. In Line 12, the algorithm the following iteration is initiated by resetting the internal training stage example epoch. Line 14 is used to estimate the next step in (look ahead estimation one single) (look ahead estimation one single) Line 19 to 27 performs the prediction for the known dataset. Line 28 to 32 provides our model performance such as RMSE, Accuracy and PMSE

4.4.1.6.1 Listing 2: LSTM implementation

```
Inputs  : GER (Time series data)
Outputs         : RMSE
1.      sizen ← length (Ger Dataset) * 0.70
2.      train ← GER [0 to size]
3.      test ← GER [size+1 to length (Ger Dataset)]
4.      set the random.seed(10)
5.      X ← training (70% of dataset)
6.      y ← training – X (30% of dataset)
7.      model = Use Sequential()
8.      model.layer .add(LSTM (neurons))
9.      model.layer.compile (loss=MSE, optimizer= RMSProp)
10.     for each in (X) do
11.             model.layer.fit(X, y)
12.             model.resume()
13.     end for
        return model
14.     ForcastGER ← model.predict(y)
        return ForcastGER
15.     MSE ← Mean Square Error(expected, ForcastGER)
16.     RMSE ← sqrt (MSE))
17.     Return RMSE
```

4.4.1.7 Model Selection

The model selection compares both the statistics and machine learning models in terms of prediction accuracy. The compare models were developed during the training phase on the known time series data. This selection is performed on the testing phase; where prediction is perform against the truth data also known as the known datasets. Assessment Metrics such as RMSE, Accuracy and PMSE were utilized to examine the model.

In our model selection K-Fold cross validation technique is utilized. The total dataset is separated into 5 folds for the year between 1981-82 and 2019-20 as indicated in Table 4.1 below. The 2^{nd} column indicate the starting year, 3^{rd} column gives the ending year and the last column give the no of year with the fold sub-sets.

Table 4.1: Training and Test data distribution for 5 folds

Fold No.	Start Year	End Year	No. of year
F1	1981	1988	8
F2	1989	1996	8
F3	1997	2004	8
F4	2005	2012	8
F5	2013	2019	7

The dataset for training and testing is scuffled so that the input and output are completely random. Thereafter the dataset is split into 5 folds as in Table 4.1. The numbers of years covered in each fold are identical from F1 to F4, however due to the data distribution variation, F5 contains 7 years. The main purpose of scuffling and splitting the dataset is to achieve randomized dataset. This also enabled to make the trained model more robust to predict unseen datasets. Selecting an accurate model after randomizing using K-fold during training generalized the model without over-fitting the training phase. The overall process of model selection is depicted in Figure 4.11.

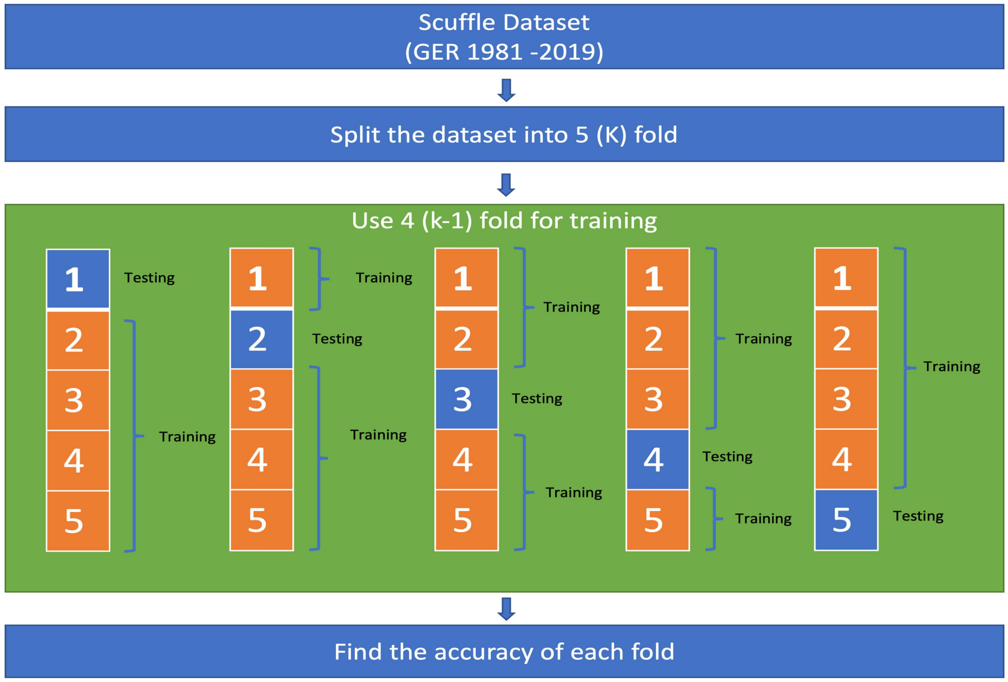

Figure 4.11: Model selection using 5-Fold

Model selection using a 5-Fold graphical representation shows how much data were used for training and testing in each data fold. In Fold -1 out of [1,2,3,4,5] datasets, the first dataset (1) is used for training while the remaining dataset [2,3,4,5]. Likewise, the training dataset is randomized from FOLD 2 to FOLD 5 by taking a section of 2, 3, 4 and 5, respectively, as represented in Figure 4.11. Data scuffling using K-FOLD is a common practice in training models as it helps generalize the model.

4.4.1.8 Forecasting

In this section, the actual forecasting of GER is performed using the selected LSTM model, as the previous section (model selection) has confirmed that LSTM is superior in prediction accuracy. A new LSTM model is built using 100% of the known GER data (1981 to 2019). In other words, all the available dataset is used for training the LSTM model. This approach is practically feasible as the comparison data in the model selection stage has suggested that LSTM-based models beat ARIMA-based models by a substantial margin. Therefore, the LSTM model is selected for performing the GER Forecasting.

4.4.2 Comparison of machine learning and forecasting

In this section we will introduce the methodology use to analyses and predict student enrolment in details. Figure 4.12 below depicts the overview of the implemented methodology.

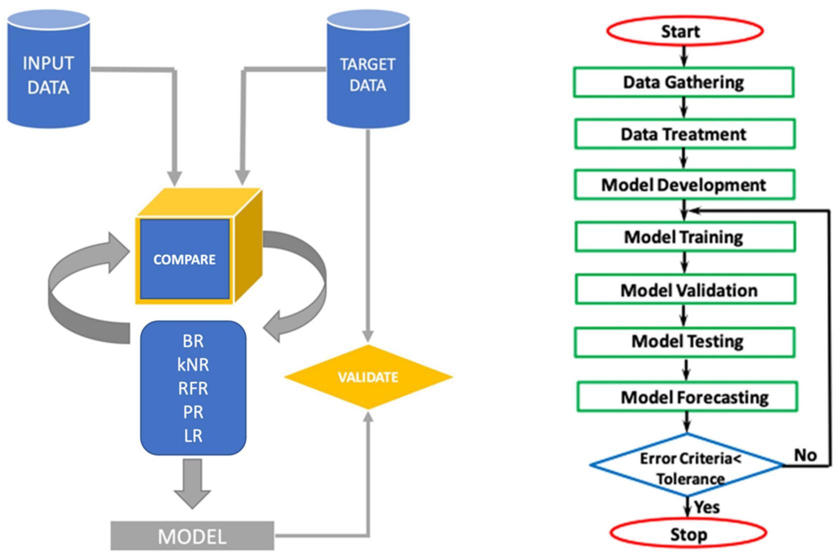

Figure 4.12: Overview of implemented methodology

4.4.2.1 Data Collection

The primary purpose of data collection is to gather useful information. In this study data were obtained student enrolment data from the yearly report of All India Survey on Higher Education (AISHE), published by the Ministry of Human Resource Development (MHRD), Govt. of India and Statistical Handbook of various years published by Dept. of Economics & Statistics, Govt. of Mizoram. The data is collected for a period of 52 years (1968 – 2019). The enrolment data is depicted in below Figure 4.13, Figure 4.14 and Figure 4.15.

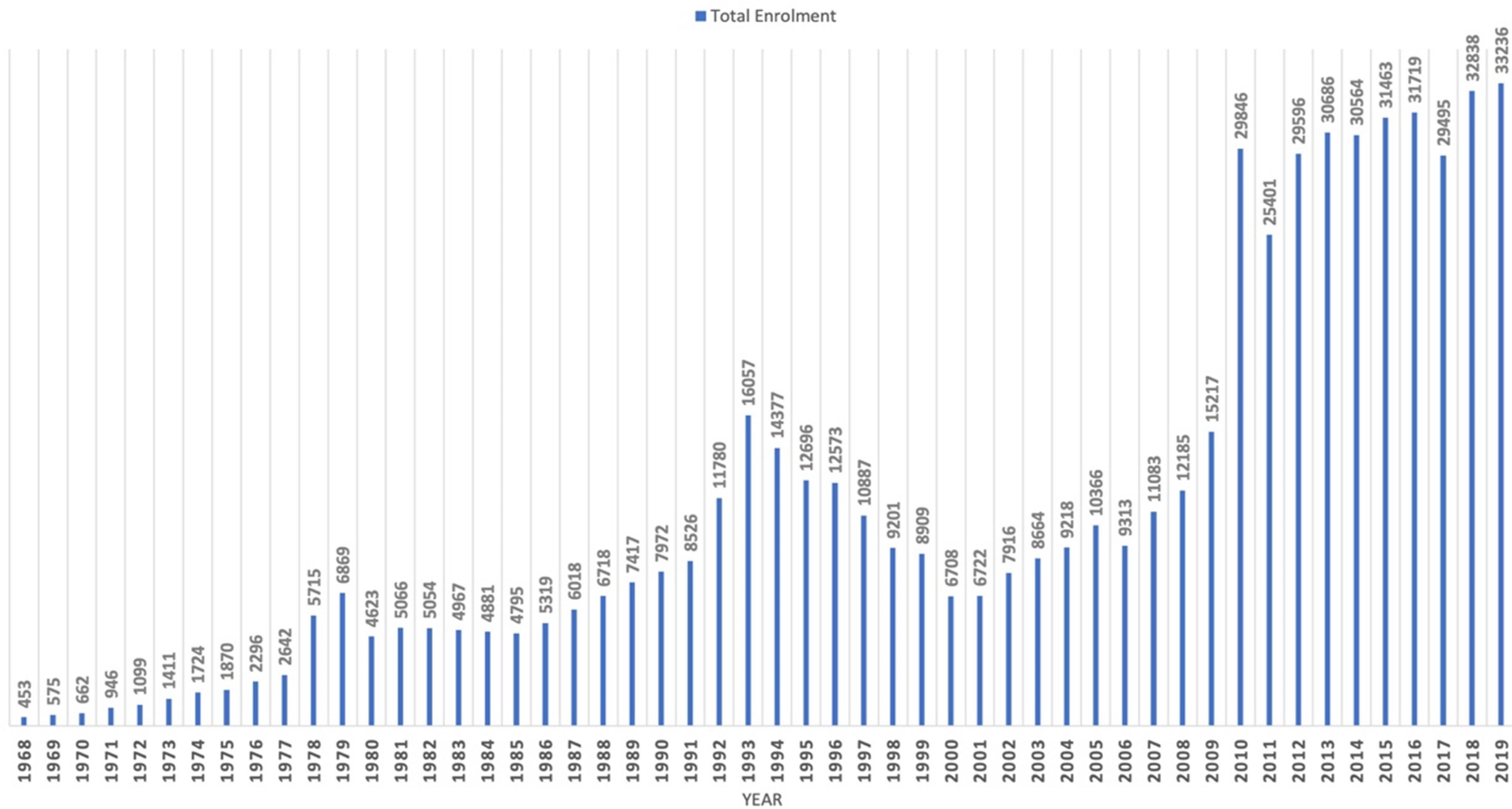

Figure 4.13: Total Student Enrolment

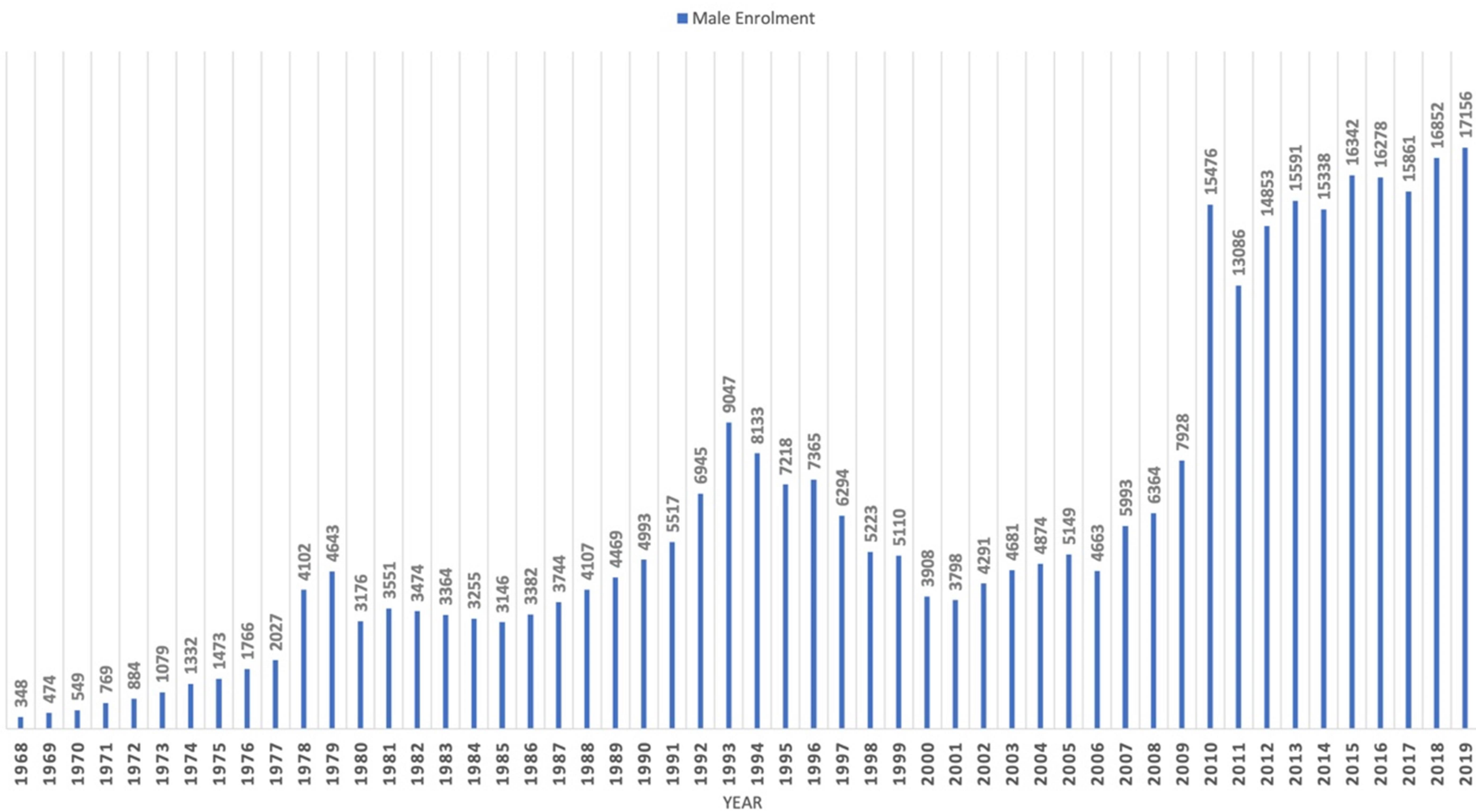

Figure 4.14: Male Student Enrolment

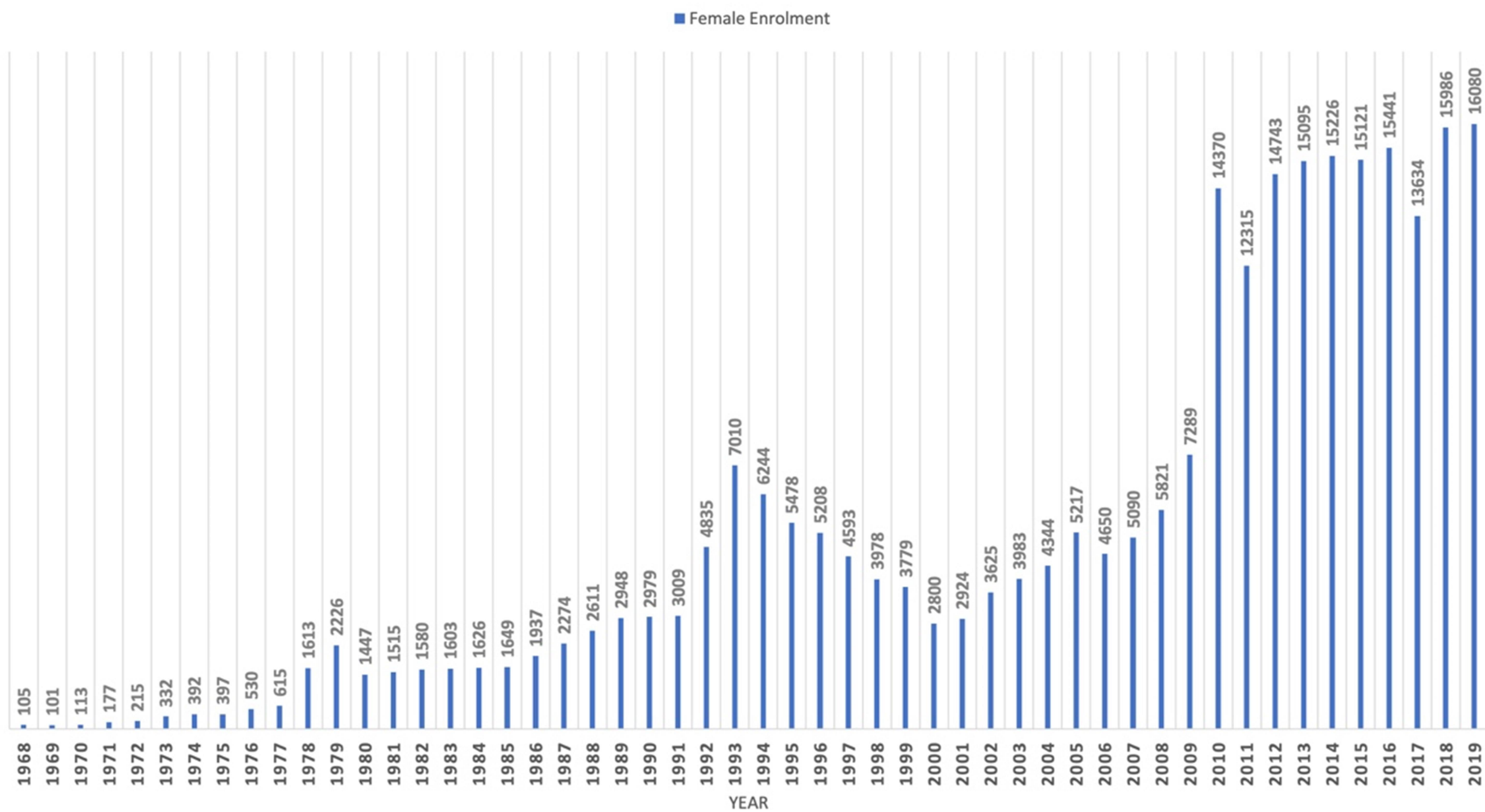

Figure 4.15: Female Student Enrolment

4.4.2.2 Data Pre-processing

The student enrolment data obtained from the data collection require pre-processing in order to use machine learning for analysis and prediction. The data is therefore categories into two classes known as the input and target data. The input data represent the past years on which the enrolment details are captured. The target data represent the parameter of the enrolment details for a particular year which corresponds to the input data. The feature of target data consists of parameters such as male enrolment, female enrolment and total enrolment of student for each year. After generating this input and target data, a machine learning algorithms are used for building model. The model is trained using 596334 student samples with the data compiled from admissions records. In order to obtain an accurate enrolment prediction, five model of regression network known as Bagging regression, KNN Regression, Random Forest Regression, Polynomial Regression and Linear Regression were analysed and compared.

4.4.2.3 Model Development

A generalized Architecture with similar setup has been implemented for the four types of neural networks. The results for this initial architecture experimentation indicate that the hyperbolic tangent activation function, momentum of 0.8, and learning rate of 0.1 achieve the best performance results and are used as the values for subsequent architecture experimentation. Next, the number of hidden layer neurons are varied from 10-70 while adjusting the learning technique to examine their impact on performance and determine the optimal architecture. In this hit and trial method a hidden node with 14 is found to be most optimal for all the architecture. The generalized parameters are given in Table 4.2 below. The performance metrics used to evaluate the performance of the competing neural network architectures include: MSE achieved, computation time, number of epochs, classification accuracy, recall, precision.

Table 4.2: Generalized parameters for Neural Network Architecture

S/N	Parameters	Values	S/N	Parameters	Values	S/N	Parameters	Values
1	Layers	3	4	Learning rate	0.1	7	Transfer function	Sigmoid

2	Input Node	3	5	Output Nodes	1	8	Epoch	1000
3	Hidden layer	2	6	Optimization Algorithm	Gradient descent	9	Momentum	0.8

In this section, we present the overall modelling process. The implementation of the ANN model can be described using the flow chart in Figure 4.12. Historical data from 1968-2019 (52 Years) of student enrolment has been gathered from the state of Mizoram as shown in Figure 4.13. Based on the gathered data, this study will develop a forecast model that predicts future enrolment for the years 2020 to 2035. Using the ANN technique, training and learning procedures are fundamental in forecasting future events. The training of feed forward networks is usually carried out in a supervised manner (Lillicrap *et. al*., 2016). With a set of data to be trained (usually extracted from the historical data), it is possible to derive an efficient forecast model. The proper selection of inputs for ANN training plays a vital role to the success of the training process. On the other hand, the learning process involves providing both input and output data, the network processes the input and compares the resulting output with desired result. The system then adjusts the weight which acts as a control for error minimization. In order to minimize error, the process is repeated until a satisfactory criterion for convergence is attained. The knowledge acquired by the ANN via the learning process is tested by applying it to a new data set that has not been used before, called the testing set. It should now be possible that the network is able to make generalizations and provide accurate result for new data. Due to insufficient information, some networks do not converge. It is also noteworthy that over-training the ANN can seriously deteriorate forecasts. Also, if the ANN is fed with redundant or inaccurate information, it may destabilize the system.

Training and learning process should be thorough in order to achieve good results. To accurately forecast, it is imperative to consider all possible factors that influence student enrolment, which is not feasible in reality (Hashim *et. al*., 2020). In this paper, different criteria were used to evaluate the accuracy of the ANN approach in forecasting student enrolment in Mizoram. They include: mean squared error

(MSE), root mean squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE).

A popular and important criterion used for performance analysis is the MSE. It is used to relay concepts of bias, precision and accuracy in statistical estimation. Here, the difference between the estimated and the actual value is used to get the error, the average of the square of the error gives an expression for MSE. The MSE criterion is expressed in equation 14.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y_i})^2 \qquad (14)$$

where y_i is the actual data and y_i bar is the forecasted data. The RMSE is a quadratic scoring rule which measures the average magnitude of the error. The RMSE criterion is expressed in Equation (15). RMSE usually provides a relatively high weight to large errors due to the fact that averaging is carried out after errors are squared. This makes this criterion an important tool when large errors are specifically undesired.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y_i})^2} \qquad (15)$$

Validation techniques are employed to tackle fundamental problems in pattern recognition (model selection and performance estimation). In this study, as popular regression model such as Bagging Regression, KNN Regression, Random Forest Regression, Polynomial Regression and Linear Regression are employed. The validation set will be used as part of training and not part of the test set. The test set will be used to evaluate how well the learning algorithm works as a whole.

4.4.2.4 Model Training

The developed model for prediction is trained using the data obtained mention in methodology I. The input training data and target data are composed of student enrolment for 52 years classified into male and female. These input parameters are feed to 5 types of regression neural network architecture known as Bagging Regression, KNN Regression, Random Forest Regression, Polynomial Regression and Linear Regression.

4.4.2.5 Model Validation

Validation techniques in machine learning are used to get the error rate of the ML model, which can be considered as close to the true error rate of the population. If the data volume is large enough to be representative of the population, you may not need the validation techniques. However, in real-world scenarios, we work with samples of data that may not be a true representative of the population. This is where validation techniques come into the picture.

In this part we analyse three validation technique and compare the output as follows:

4.4.2.5.1 K-Fold Cross-Validation:

In this technique, k-1 folds are used for training and the remaining one is used for testing as shown in the Figure 4.16 given below.

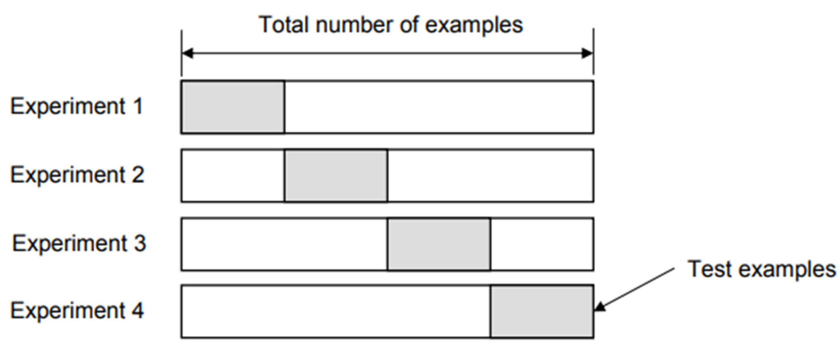

Figure 4.16: K-Fold validation

The advantage is that entire data is used for training and testing. The error rate of the model is average of the error rate of each iteration. This technique can also be called a form the repeated hold-out method. The error rate could be improved by using stratification technique.

4.4.2.5.2 Leave-One-Out Cross-Validation (LOOCV)

In this technique, all of the data except one record is used for training and one record is used for testing. This process is repeated for N times if there are N records. The advantage is that entire data is used for training and testing. The error rate of the model is average of the error rate of each iteration. The following Figure 4.17

represents the LOOCV validation technique.

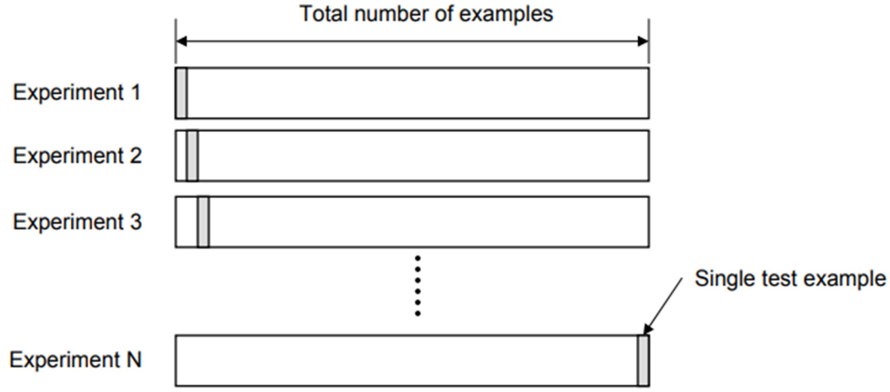

Figure 4.17: LOOCV validation

4.4.2.5.3 Random Subsampling

In this technique, multiple sets of data are randomly chosen from the dataset and combined to form a test dataset. The remaining data forms the training dataset. The following diagram represents the random subsampling validation technique. The error rate of the model is the average of the error rate of each iteration. Figure 4.18 represent random subsampling technique.

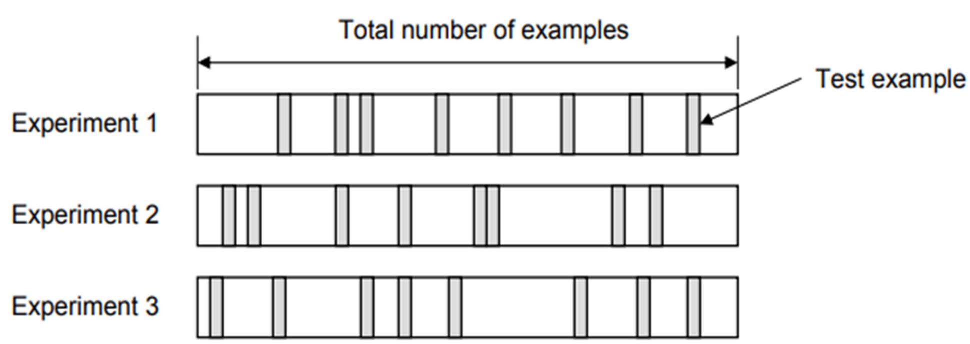

Figure 4.18: Random Subsampling

4.4.2.6 Model Forecasting

The objective of this model forecasting is to estimate the value of an unknown variable which is the student enrolment for the coming years 2020-2035. A trend exists when a series increases, decreases, or remains at a constant level with respect to time. Therefore, the time is taken as an input feature. However, in our scenario, Cycles seasons that do not occur at a fixed rate or interval is the behaviour of our enrolment trend generated from the past known data. They do not repeat at regular time intervals, However, we should remember that these variables need to have definite patterns. Therefore, to identify this pattern Neural Network is taken to handle. Both time series components and features are key to interpreting the behaviour of our prediction.

4.5 Results and Discussion

This section is categories into two sub section, in the first sub section the result finding of both the comparison between statistical methods and machine learning are elaborated in details. In the second sub section the comparative results of various machine learning model are discuss.

4.5.1 Machine learning and statistical method

Table 4.3: ARIMA Performance

Fold No.	RMSE	Accuracy %
1	0.478	87
2	0.323	86
3	0.451	82
4	0.359	87
5	0.212	91
Average	0.365	86.6

Table 4.4: LSTM Performance

Fold No.	RMSE	Accuracy %
1	0.321	91
2	0.218	83
3	0.193	94
4	0.231	97
5	0.199	96
Average	0.2324	92.2

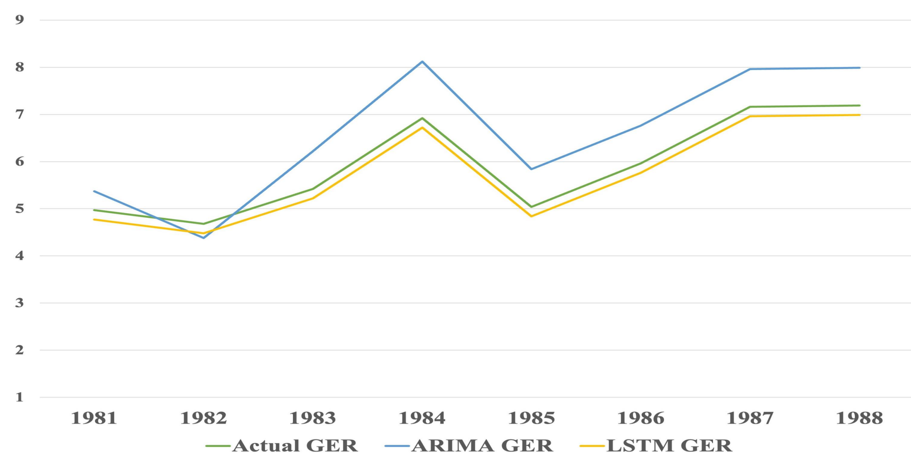

Figure 4.19: Fold 1– Performance of ARIMA and LSTM

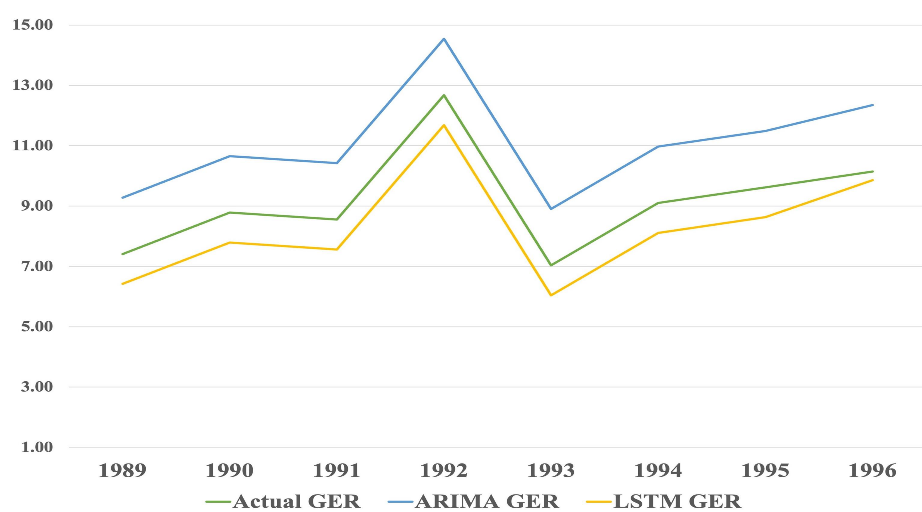

Figure 4.20: Fold 2 – Performance of ARIMA and LSTM

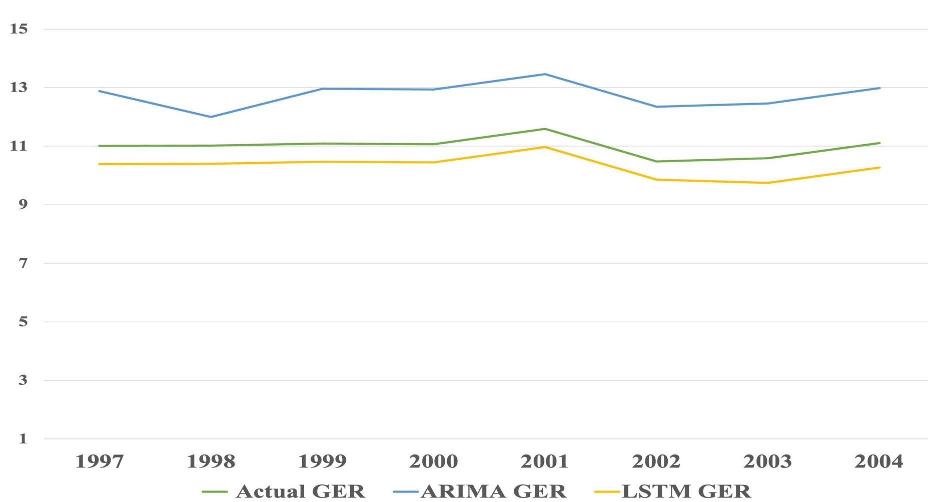

Figure 4.21: Fold 3 – Performance of ARIMA and LSTM prediction

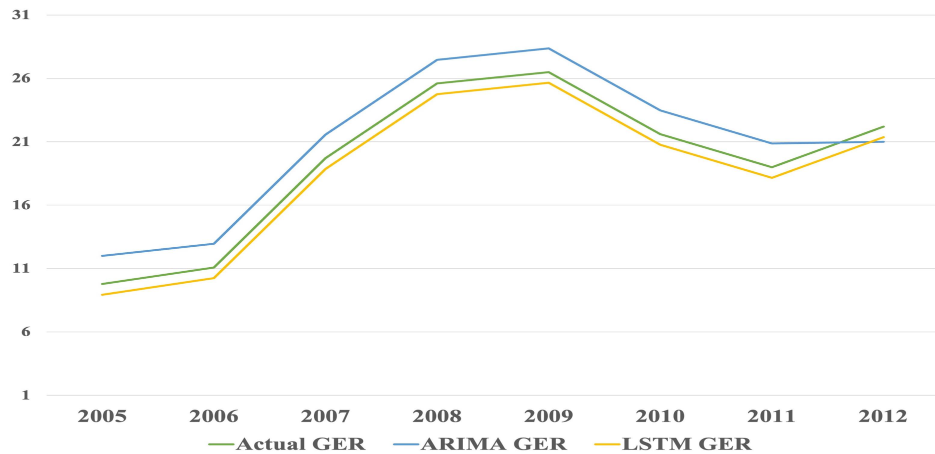

Figure 4.22: Fold 4 – Performance of ARIMA and LSTM

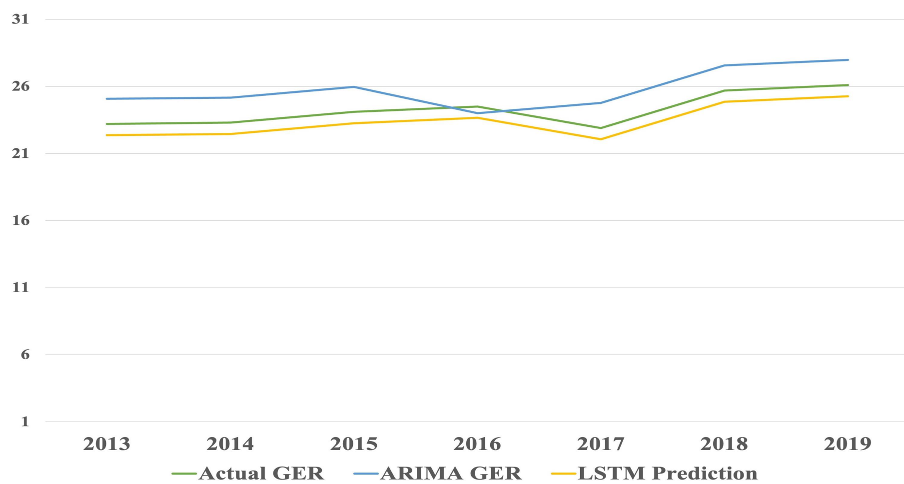

Figure 4.23: Fold 5 – Performance of ARIMA and LSTM

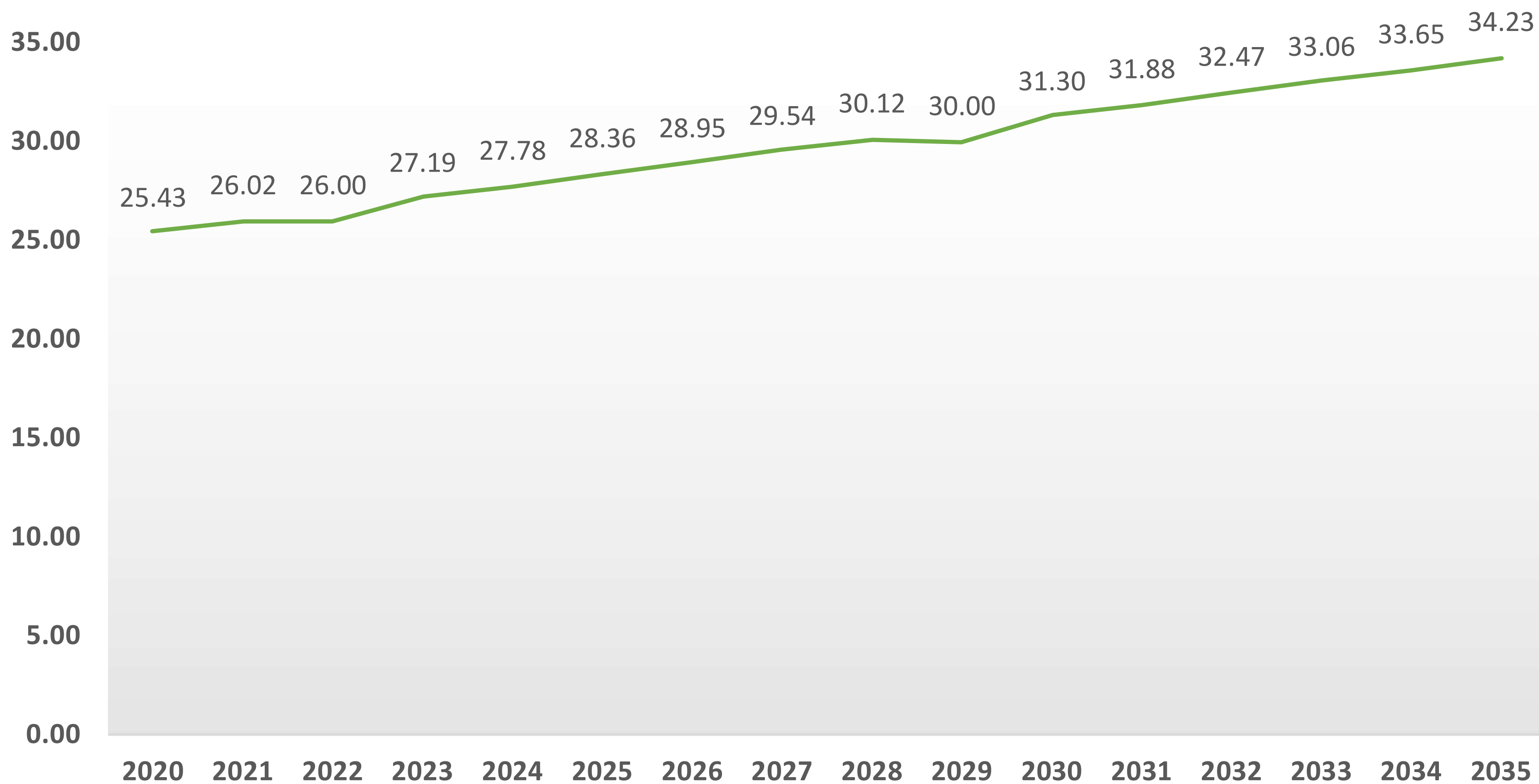

Figure 4.24: GER forecasting using LSTM (2020 - 2035)

Table 4.5: Enrolment at various levels in Mizoram (2019-20)

District	Ph.D	M.Phil	PG	UG	PG Diploma	Diploma	Certificate	Integrated	Total
Aizawl	867	152	4221	19745	203	984	463	73	26708
Champhai				913		92			1005
Kolasib				544		87			631
Lawngtlai				1183		82			1265
Lunglei				1944		517			2461
Mamit				181		57			238
Saiha				359		64			423
Serchhip				427		78			505
Mizoram	867	152	4221	25296	203	1961	463	73	33236

The study's findings indicate that LSTM is superior to ARIMA in forecasting time series data. A comparison of Table 4.3 and Table 4.4 reveals that the LSTM-based algorithm predicts more accurately than ARIMA by 5.6 percent and that RMSE reduces in LSTM by 0.1322. Furthermore, GER prediction performance against the actual known datasets from Figure 4.19 to Figure 4.23 demonstrates that LSTM is significantly closer to the actual GER value, which favours LSTM as compared to ARIMA. These experimental findings direct the research to select the LSTM model for forecasting future GER data. The GER forecasted value (2020-2035) using the selected LSTM is plotted in Figure 4.25. In this forecasting, GER increment is observed 14 times, and decrement is observed 2 times (2022 and 2029). This decrement in GER indicates that the number of student enrolment for these years is expected to be lower than the previous year. The average GER variation during these forecasting periods is 0.59. The forecasted GER mean and standard deviation are 31.73 and 12.22, respectively.

Improvement in accuracy and RMSE were not observed when increasing the number of epoch during the training phase. The exceptional performance observed through LSTM-based methods is due to the "iterative" optimization technique utilized in these approaches to obtain the best outcomes. The NEP 2020 aims to increase the GER to 50 % by 2035. The forecasting GER using our selected model (LSTM) for 2035 is 34.23 % which is lower than the NEP target by 15.8%. Considering the latest enrolment data in Table 4.5, the number of student enrolment from UG to PG drop significantly by 83.3 %. This is clear indication in shortage of seats and limited number of institutions for PG courses. The policymaker may consider increasing the number of seats and institution to achieved NEP 2020 target.

4.5.2 Machine learning architecture

The first step in evaluating the results is to compare the machine learning methods in terms of their prediction performance. Table 4.6 below shows the prediction results of the five machine learning architecture for the raw data set, which corresponded to the pass yearly student enrolment categories into male, female and total. Each of the regression model are compare against Mean Absolute

Percentage Error and R^2. The result listed in Table 4.6 clearly identifies that, Bagging Regression model is the most effective architecture among the others.

Table 4.6: ANN architecture comparison for the known dataset

REGRESION MODEL	MAPE	R^2
Bagging Regression	0.054	0.98
KNN Regression	0.083	0.97
Random Forest Regression	0.622	0.94
Polynomial Regression	0.434	0.87
Linear Regression	0.824	0.74

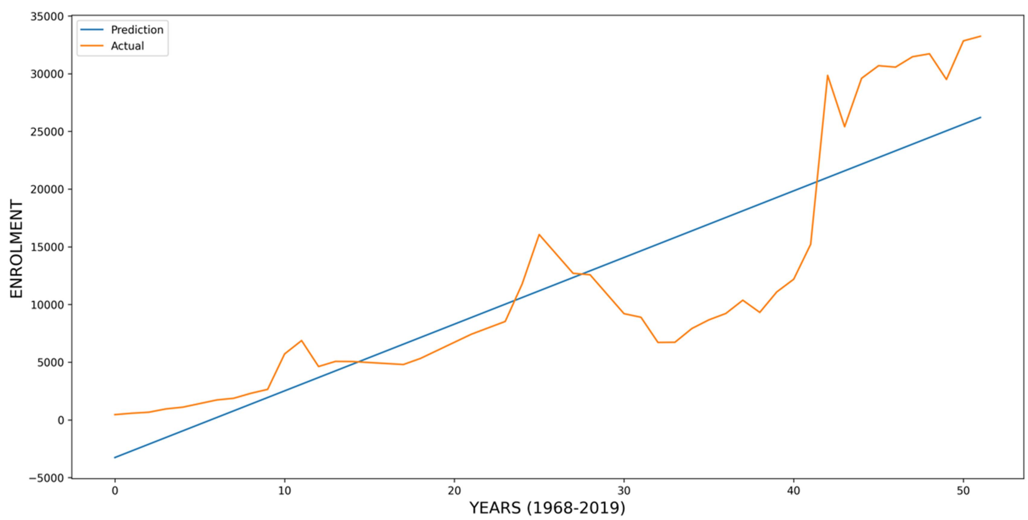

Figure 4.25: Performance of Linear Regression

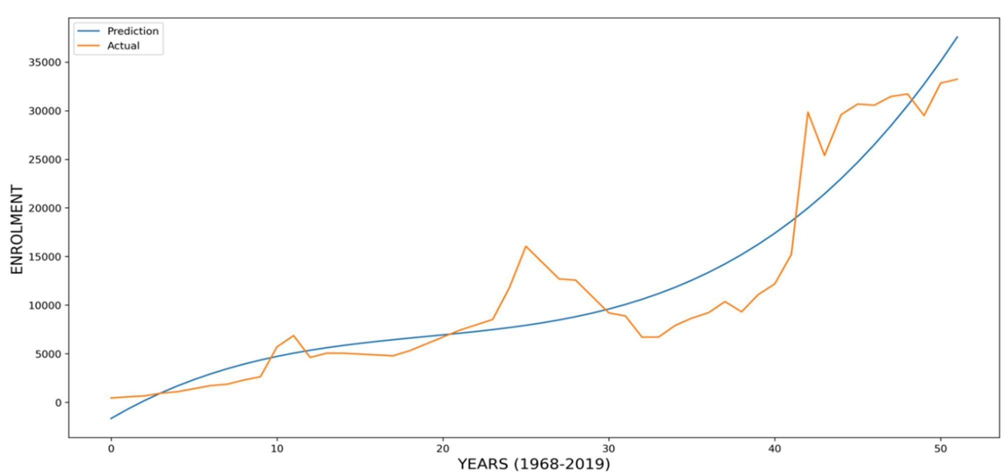

Figure 4.26: Performance of Polynomial Regression

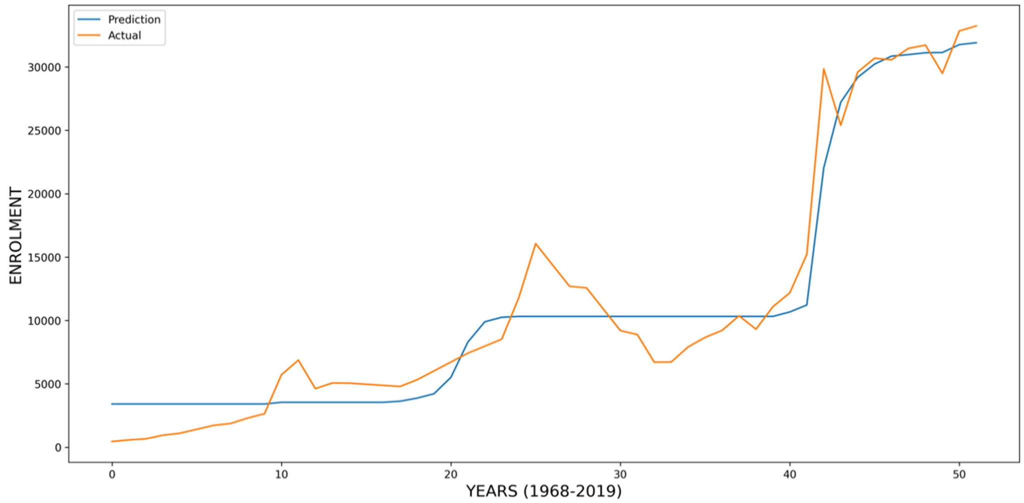

Figure 4.27: Performance of Random Forest

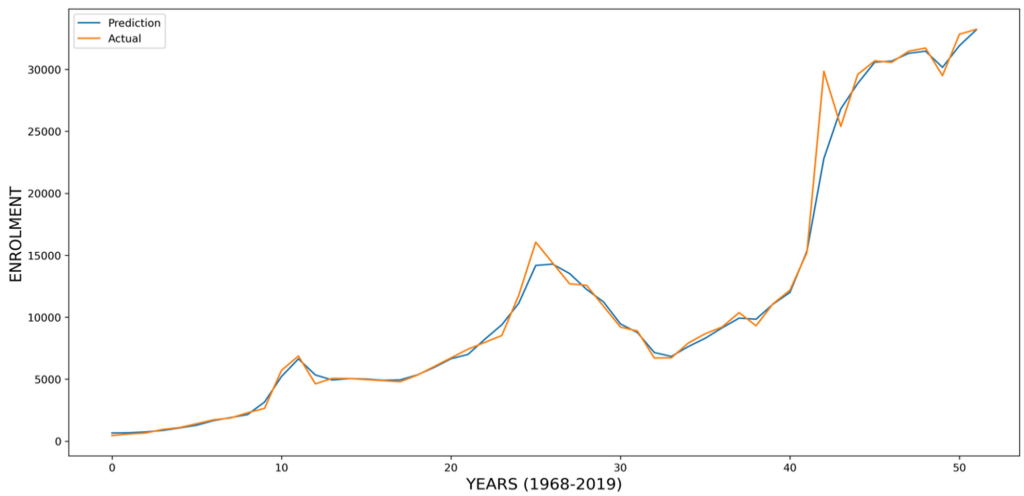

Figure 4.28: Performance of Bagging Regression Model

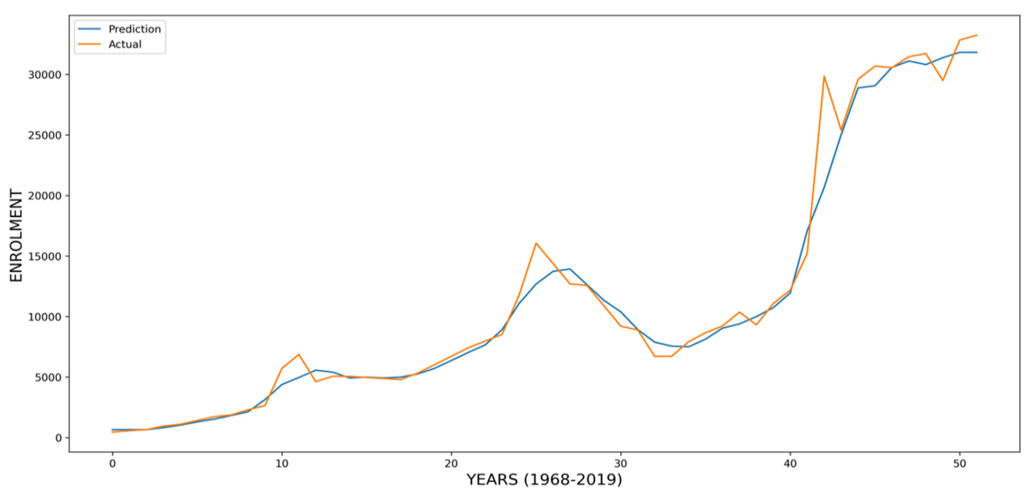

Figure 4.29: Performance of KNN Regression

The above Table 4.6 gives the model performance evaluation. Lower MAE and MSE indicate the model with higher accuracy. Likewise higher value of R^2 represent model with higher accuracy.

The comparison for five types of machine learning architecture is given in Figure 4.30 below:

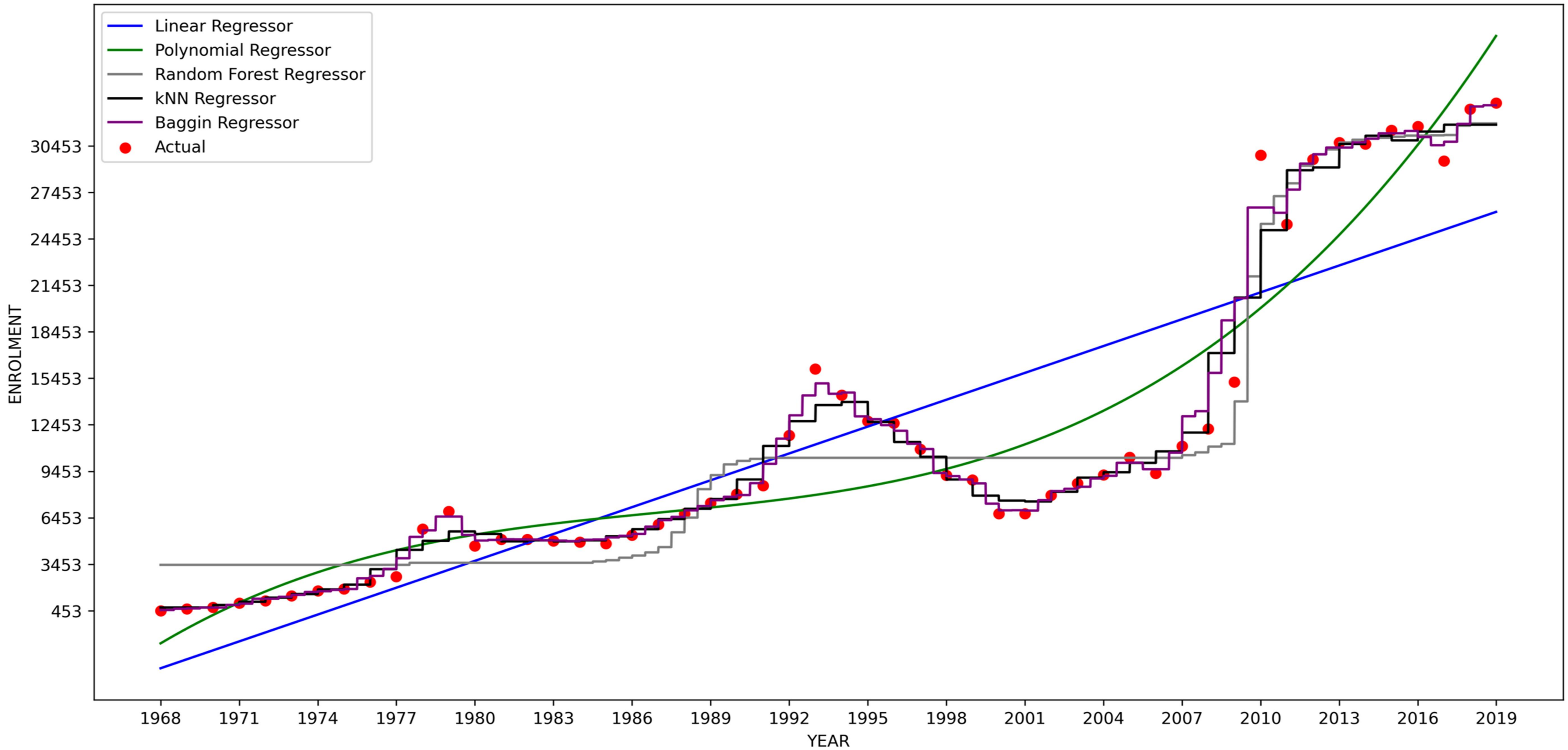

Figure 4.30: Performance of five types of machine learning architecture

From the analysed result set obtained from the above Table 4.6, Figure 4.25, Figure 4.26, Figure 4.27, Figure 4.28 and Figure 4.29, it is clear to draw a decisive conclusion that Bagging Regression model is better as compare with the others architecture for predicting the unknown student enrolment for the year 2020 to 2035. The enrolment forecasting for 2020 and 2035 is given in Figure 4.31 below:

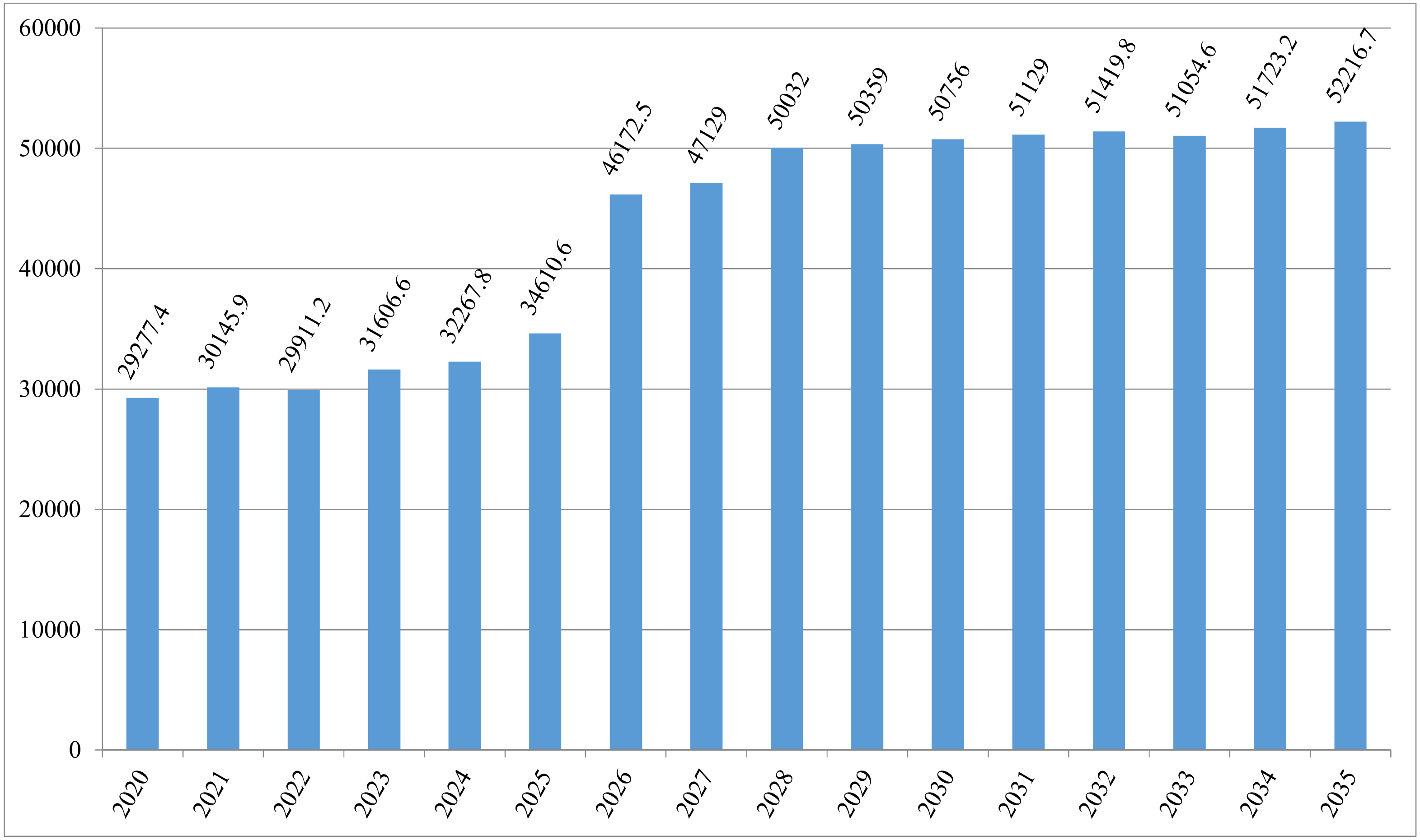

Figure 4.31: Enrolment forecasting using Bagging Regression (2020 - 2035)

4.6 Conclusion

This Chapter conducts a comparative analysis of two state-of-the art forecasting techniques: ARIMA and LSTM. In this process, time series data have been pre-processed by transforming the yearly enrolment dataset into the equivalent GER. This GER data is then split into 5 folds, so as to train each fold separately. The results of each fold were examined and compared to the known dataset. For future GER forecasting, the model selection phase chooses the LSTM, which is one type of machine learning model. Therefore, in the second phase, the study is redirected to performed more in-depth performance analysis of different architecture of machine learning.

In this second phase, it is determined that the best algorithm for predicting student enrolment is Bagging Regression model. The ideal perspective for building the algorithm is to use information for all of the known student enrolment for the past 52 years obtained from AISHE report and Statistical data of Mizoram. The predictive capacity of this algorithm is the best of the alternatives evaluated.

In addition to yielding the best MAPE and R^2 value, the Bagging Regression showed a closest plot to the actual value while performing prediction. It is found that both the ability to correctly detect enrolment increase over time. The results also suggest that, to train the student enrolment prediction algorithm, it is convenient to exclude categories of the student such as SC, ST and GENERAL student as it added noise in the dataset.

This research finding provides policymakers with insight into future academic enrolment. Future research may be carried out by examining the status and reasons impacting student dropout at different levels of education since it has a direct and indirect influence on higher education (GER).

CHAPTER 5

SUMMARY AND CONCLUSION

This chapter summarizes the previous chapters and provides certain concluding remarks based on research findings. The main goal of this study is to ascertain the current state of enrolment in higher education and to make recommendations for increasing enrolment. The current research is centred around higher education GER. Its primary objective is to emphasise on importance of access to higher education.

The study will primarily concentrate on the state of Mizoram. Data for the study has been gathered from a number of sources, including AISHE, UDISE, NSSO, the Ministry of Education of the Government of India, the Department of School Education of the Government of Mizoram, and the Department of Economics and Statistics of the Government of Mizoram. After that, the gathered information is examined making use of the proper mathematical and statistical techniques.

The results in this study is data driven. In the first chapter, a general introduction to the study, its significance and primary purpose, and the conceptual framework of the research investigation are presented. Following a review of the trend in higher education enrolment in developed countries, higher education enrolment in India as well as other indicators are discussed. Finally, the chapter focuses on higher education in the state of Mizoram.

A comparison of the state of higher education in India in 1947 and 2019 was done to discuss the expansion of higher education in India since the country's independence. Despite having one of the largest higher education systems in the world, the GER for higher education in India is still significantly lower than the average for a number of other developing nations as well as developed countries. The number of institutions, GER and PTR of Indian States and Union Territories in 2019 were discussed. In addition, the classification of institutions accredited by NAAC as on June 21, 2022 according to grade, as well as the number of universities and colleges in each state that are accredited by NAAC, were both discussed.

This study aims to determine the status of higher education in Mizoram. Utilizing statistics and various forms of machine learning, it seeks to evaluate the number of students enrolled as well as GER of higher education in Mizoram. This chapter contains a review of a variety of literature that is relevant to the present study. The majority of research focuses on GER of higher education, but the current study also used statistical and machine learning technique to provide systematic and scientific knowledge of the problem and predict potential outcomes. The purpose of the study was to look at GER in higher education. Statistical and machine learning techniques are used to give a systematic and scientific understanding of the problem and predict GER to know whether Government of India's target of 50% is achievable or not.

The second chapter examines the impact of student dropout on GER. In this analysis, data on student dropout at lower (School) level is gathered from official and independent databases, government archives, journals, publications, and the internet. Mathematical and statistical analysis is used to analyse the data. The dropout dataset is prepared from nineteen years data from 2001 to 2019. The dataset is then merged with the historical enrolment by adding each dropout to its appropriate enrolment. Higher education GER without dropout is then calculated by estimated population of age group 18-23 years. This dataset demonstrates how dropouts at higher education levels impact the GER. This study emphasizes the importance of student retention during the course. The data has been analyzed by employing analytical techniques. Time-series enrolment data with negligible dropout is used as an input to create a model in these techniques. The model construction procedure is initiated by a data-stationary test. This test is carried out by running three procedures: the decomposition plot, the mean variance test, and the augmented Dickey-Fuller test. The test techniques demonstrate the dataset is stable, which means no differentiation is necessary to be conducted in order to be utilized as an input to the proposed model ARIMA. The model training is accomplished by utilizing all the available GER of 19-year datasets. During this training, RMSE is considered a model accuracy indicator. The model created in this training phase is utilized to forecast the future GER over a period of six years. This forecasted assessment demonstrates the demands of minimizing dropout and the significance of student retention in order to boost GER.

From the analytical study of data on student dropout the following points are concluded.

- Among different levels of school education it is discovered that the overall GER is highest at the Primary School level, with scores ranging from 118.89 to 158.89 over the past ten years.

- This score then gradually decreased upward in the Upper Primary School level, with scores ranging from 98.06 to 115.61.

- Finally, this score decreased once more in the Secondary School level, with scores ranging from 81.9 to 94.68, and it dropped dramatically in the Higher Secondary School level, with scores falling in the range of 44 to 61.3.

- The GER pattern is the same for both the girls and the boys as well as for the overall category. From the secondary level on up to the higher secondary level, it is observed that the GER has been steadily falling for some time now.

In the second part of chapter 2, implication of student dropout at different level of school education in higher education GER were studied in details. In this studies, growth of schools with respect to student enrolment (primary, middle, high school and higher secondary school) were first identified. GER for Boys, Girls and Total were derived from this enrolment. The mean GER was calculated. Thereafter, dropout analyses were performed at each level for all 8 districts in Mizoram. After identifying the mean drop out in each level of school education, a new GER is formulated by recalculating the enrolment without considering the number of dropouts. The new GER without dropout is further analyze and forecast using statistical methods.

The statistical method employed in these studies is known as ARIMA, where the new enrolment GER without dropout forms a new dataset. This new dataset can be interpreted as a timeseries data, therefore; in initial stage of this analyses, 3 types of stationary such as decomposition plot, mean and variance and augmented Dickey-Fuller test were performed.

The decomposition plot time series decomposition is a process of deconstructing a time series into the following components:

- Trend — general movement over time
- Seasonal — behaviors captured in individual seasonal periods
- Residual — everything not captured by trend and seasonal components

This technique is most often used for analyzing historical time series data. It's also often used for forecasting. Modeling trend and seasonality at once might be a too difficult task, so tackling components individually serve a better approach.

The mean and variance of the time series dataset give a brief description on how the data changes over a period of time. These changes indicate how time invariant is the dataset. A constant mean and variance over a period of time is a confirmation for stationarity of the data. The dataset in this study is divided into two independent sets. The mean and variance were calculated for each set, and the finding shows the variation in each set correlated to the stationary of the data.

The final test employs in this study is known as Augmented Dickey–Fuller test. This test is an extension of the dickey-fuller test, which removes autocorrelation from the series and then tests similar to the procedure of the dickey-fuller test. The augmented dickey fuller test works on the statistic, which gives a negative number and rejection of the hypothesis depends on that negative number; the more negative magnitude of the number represents the confidence for stationary dataset. All these three tests favour our timeseries dataset to be stationary.

ARIMA prediction using all the known dataset is performed by taking p, d, q value as (1, 0, 0). The Mean and RMSE for the above prediction are 50.86 and 1.82 respectively. The prediction model is saved and forecasting using the same is carried out for the next 6 years (2020-2025) where enrolment is unknown.

It is found that student dropout at school education has a major impact on higher education GER. The impact of student dropout on higher education GER is demonstrated by the fact that, if dropout rates at all educational levels are eliminated, higher education GER will be roughly 43.8 percent, which is 17.7 percent higher than it is right now.

In the third chapter, the GER of higher education for each district has been calculated. Higher education enrolment and GER throughout the states are analyzed according to educational level, district, gender, and mode of study. Furthermore, the relationship between the GER of higher education and the GER of other educational levels in the schooling system is investigated. The following results are obtained.

- The state has three Universities, thirty five colleges, and sixteen stand-alone institutions.
- Except for a sharp drop in 2014–15, Ph.D. enrolment goes up steadily from 2010 to 2019.
- M.Phil. enrolment in Mizoram does not follow a regular pattern. The lowest enrolment was in 2011–12, and the highest enrolment was in 2018–19. But it is clear that the number of M.Phil. students has grown overall from 2010 to 2019.

- Postgraduate enrolment seems to be slowly going up during the study period, with the lowest number reported for the academic year 2011–12. But the enrolment trend from 2016 to 2019 shows an uneven pattern.
- From 2010 to 2019, the number of undergraduate students is going up five times and going down four times. During the time of the study, the enrolment trend was also not steady here. It is seen that there is not much increase in Post Graduate Diploma enrolment from 2010 to 2018. The academic year 2018-19 recorded a steep rise in enrolment but again receded the following year.
- Over the course of the study period, enrolment in the diploma goes up three times and down six times, with no clear pattern.
- From 2010 to 2017, not many students enrolled for Certificate courses. But starting in 2018, there were a lot more people taking these certificate courses.
- It is observed that Mizoram just saw its first integrated enrolment in 2014–15, however there is a growing trend till 2018–19. During the past ten years (2010–2019), higher education enrolment in Mizoram increased from 29846 in 2010 to 33236 in 2019, which is found to be an increase of 11.36 percent.
- The ten-year analysis of student enrolment reveals that undergraduate studies consistently have the highest proportion of students, ranging from 73.32 percent in 2015–16 to 83.39 percent in 2012–13, followed by postgraduate studies, which range from 10.11 percent in 2011–12 to 13.0 percent in 2016–17. The overall enrolment composition of other studies, including Ph.D., M.Phil., PG Diploma, Diploma, and Integrated Courses, was similar between 2012–13 and 2015–16, ranging from 5.37 percent to 14.67 percent.
- From the academic year 2010-2011 to the academic year 2019-2020, Mizoram's GER in higher education barely increased by 4.5%.
- According to the findings of the above analysis, being the state capital Aizawl has the highest GER of all the districts in Mizoram. There could be a great number of reasons for this, but one of the most important ones could be that

people moved to Aizawl from other districts because it is the state capital and has superior amenities.

- Mamit district, on the other hand, has the lowest GER of 1.1 in 2010-11 and 3.2 in 2019-20, which is a very low number in comparison to the whole state GER, which ranged from 19.0 to 26.1 over the course of ten years.
- Comparing enrolment in higher education between men and women in the state of Mizoram from 2010 to 2019, it is good to note that gender disparities are small.
- Distance education enrolment fluctuates between 5,044 in 2017-18 and 7,995 in 2013-14. The majority of distant education students are enrolled in undergraduate and graduate programmes.
- The degrees of relationship between GER of higher education and GER at different levels of education are calculated by Pearson correlation coefficient by comparing the overall GER of higher education to the overall GER of other educational levels. It has been found that the correlation between higher education and higher secondary school has a value of 0.863, making it the strongest one discovered. This value indicates that as the GER for higher secondary school increases, the GER for higher education as a whole will also increase.
- The GER for elementary schools, on the other hand, has the weakest correlation coefficient of all the GERs with the GER for higher education; this coefficient's value of 0.088 indicates that the GER for elementary schools has the least impact on the GER for higher education.

In the fourth chapter, the research is carried out in two stages. In the first phase, model evaluation and comparison utilizing statistical approaches (predictive analytics) and machine learning were undertaken. The forecast accuracy for the known datasets (enrolment historical data) is examined in this comparison. The enrolment data for the years 1981 to 2009 is taken from the Department of Economics and Statistics, Govt. of Mizoram, and the year 2010–2019 is from the AISHE report produced by the Ministry of Education, Govt. of India. In all, 39 years of student enrolment data have been studied in this study. Data transformation is performed in the data pre-processing step by transforming the year-wise student enrolment data to its equivalent GER, which is a primary indication for enrolment status. The GER reflects the number of students enrolled as a proportion of the population aged 18 to 23. The GER created by the pre-processing stage is an input to ARIMA and LSTM for both training and testing. The input datasets are separated into training and testing at 70% and 30%, respectively. The training phase is validated using 5-fold cross-validation, as this phase is the major signal for model selection. The model selected in this stage will be utilized in the testing phase for projecting GER for the years 2020–2035. This training and testing stage is crucial, as it is the key determining element for how effectively our model can estimate the future GER. Learning weight, bias, and epoch for the LSTM model are the parameters that fluctuate throughout the training stage. Similarly, the auto regression and moving average functions are critical in ARIMA for determining the model's correctness. According to this comparative accuracy, either a machine-learning or statistical model will be picked. A five-fold validation and accuracy obtained from the confusion matrix are applied. Higher-accurate models and lower root mean square error (RMSE) are the primary considerations for model selection. The findings of this experiment benefited the machine learning model, as it beat statistical techniques by an average of 0.132% and 5.6% in both RMSE and accuracy. Therefore, machine learning is used for future enrolment prediction.

The second portion of Chapter four analyzes five types of machine learning models, which are precisely known as regression models. These regression models explored in this work are bagging regression, KNN regression, random forest regression, polynomial regression, and linear regression. MAPE and R^2 were utilized to validate our model. These data demonstrated bagging regression to be more superior in prediction accuracy as compared with the others, yielding 0.054 and 0.98 values for MAPE and R^2, respectively. Therefore, future enrolment forecasting (2020–2035) is undertaken using bagging regression.

In conclusion, the problem of low enrolment in higher education is strongly associated with high school and higher secondary enrolment, as shown by a correlation analysis of GER in higher education with GER at other educational levels. Thus, progress at various tiers needs to be coordinated for the best outcome. There is a striking correlation between GER and dropout rates, with the former occurring at a much lower rate at the elementary level while the latter peaks at the secondary and post-secondary levels. Future researchers can conducted on issues such as student dropout at the higher secondary level, which is the direct feeder of enrolment in higher education.

ABSTRACT

MATHEMATICAL AND STATISTICAL ANALYSIS
OF HIGHER EDUCATION IN MIZORAM

AN ABSTRACT SUBMITTED IN PARTIAL
FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

DAVID ROSANGLIANA

MZU REGISTRATION NUMBER : 62 of 2011
PH.D. REGISTRATION NUMBER : MZU/Ph.D./742 of 11.05.2015

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
SCHOOL OF PHYSICAL SCIENCES
JUNE, 2023

ABSTRACT

MATHEMATICAL AND STATISTICAL ANALYSIS
OF HIGHER EDUCATION IN MIZORAM

BY

David Rosangliana
Department of Mathematics and Computer Science

Supervisor: Prof. Jamal Hussain

Submitted

In partial fulfillment of the requirement of the Degree of Doctor of Philosophy in Mathematics of Mizoram University, Aizawl

ABSTRACT

Enrolment is one of the most significant indicators of higher education. The primary purpose of the study is to assess the current state of enrolment in higher education and to make recommendations for improving that condition. In order to provide systematic and scientific knowledge of the problem and to predict probable consequences, the work in this study is data-driven and makes use of statistical and machine learning technique. The study's conclusions would be useful for developing and implementing policy.

In view of the above, the current study, titled "Mathematical and Statistical Analysis of Higher Education in Mizoram" has the following objectives:

- to investigate student dropout at various stages in school.
- study the problems of low GER in higher education and propose measures to increase it.
- to develop a mathematical model of GER.

The current study is focused on higher education GER and aims to highlight the significance of access to higher education in Mizoram. Data for the study has been gathered from a number of sources, including AISHE, UDISE, NSSO, the Ministry of Education of the Government of India, the Department of School Education of the Government of Mizoram, and the Department of Economics and Statistics of the Government of Mizoram. Then, using appropriate mathematical and statistical methods, the acquired data is analyzed.

The first chapter provides an overview of the research, outlining its background, significance, and primary aim, as well as outlining the conceptual framework of the study. Enrolment in higher education is examined in the developed world first, then in India, and finally in Mizoram.

The second chapter examines student dropout and GER, assuming zero dropout. This analysis uses data from official and independent databases, government archives, journals and publications. The dropout dataset spans 2001–2002 to 2019–2020. Each dropout is added to the appropriate historical enrolment to merge the datasets. Higher education GER without dropout is then calculated by incorporating age population estimated from 2011 census data. This creates a new GER dataset with zero student dropouts. The model construction procedure is initiated by a data-stationary test. This test uses the decomposition plot, mean variance test, and augmented Dickey-Fuller test. The tests show the dataset is stable, so no differentiation is needed to use it as an input to ARIMA. Model training uses 19-year GER datasets. RMSE is an accuracy indicator during training. The model created in this training phase forecasts GER over six years. This predicted assessment shows the importance of reducing dropout and retaining students to boost GER.

In the third chapter, higher education enrolment and GER are evaluated by level, district, gender, and study style. The association between higher education GER and other educational levels is also studied. A correlation analysis of GER in higher education with GER at other levels shows that poor enrolment in higher education is substantially linked to high school and higher secondary enrolment. The GER of higher education in Mizoram barely increased by 4.5% between 2010–2011 and 2019–2020. Aizawl and Mamit are Mizoram's best and worst districts, respectively, for enrolment and GER throughout the study period.

The fourth chapter forecasts GER using data. This study's approach has two parts. This section compares machine learning and statistics. LSTM for machine learning and ARIMA for statistics are used. This work pre-processes student enrolment data to derive GER. GER analysis and predictions are performed using the state-of-the-art models ARIMA and LSTM. This study uses data from the AISHE annual reports produced by the Indian Ministry of Education and the Mizoram Department of Economics and Statistics.The comparative findings of ARIMA and LSTM indicated that

LSTM out-perform ARIMA by an average of 0.132 and 5.6 % in both RMSE and Accuracy, respectively. These results imply machine learning is more superior in comparison with statistical approaches under the present study.

The second phase of chapter 4 studies machine learning architecture. This comparative research includes five model of regression network namely Bagging Regression, KNN Regression, Random Forest Regression, Polynomial Regression and Linear Regression. These architectures are trained with the higher education enrolment dataset. This trained model was evaluated by comparing MAPE and R^2. This research shows that Bagging Regression performs best for projecting future enrolment. This study emphasizes the superiority of machine-learning based algorithms on time-series data.

Chapter 5 is summary and conclusion.

Finally, list of references is given at the end.

BRIEF BIO-DATA OF THE CANDIDATE

1. Name : David Rosangliana
2. Father's Name : H. Lalchhanhima
3. Mother's Name : P.C. Lalramliani
4. Date of Birth : 29.09.1983
5. Nationality : Indian
6. Gender : Male
7. Marital Status : Married
8. Permanent Address : LK/B-77/1, Kanan Veng, Aizawl, Mizoram – 796009
9. Email : drosangliana@gmail.com
10. Phone : 9862220289
11. Present post held : Assistant Professor, Department of Mathematics, Government Zirtiri Residential Science College
12. Academic Records:

Examination	Name of the board/ university	Year of passing	Percentage of marks	Division/ class/ grade	Subject
HSLC	MBSE	1996 (Dec)	68.3	I	Eng, Sc., Maths, SS, Mizo
AISSCE	CBSE	1999	56	II	Phy,Chem,Math, Computer Sc., English
B.Sc. Mathematics	MZU	2002	50.5	II	Phy,Chem,Math
M.Sc. Mathematics	NEHU	2007	56.25	II	Mathematics
CSIR UGC NET	CSIR-UGC	2006 (Dec)		LS	Mathematical Sciences

Date : 06.06.2023

Place : Aizawl (DAVID ROSANGLIANA)

LIST OF CONFERENCE / SEMINAR / WORKSHOP ATTENDED

1. Attended "NORTH-EAST ISI-MZU WINTER SCHOOL ON Algorithms with special focus on Graphs" organized jointly by Advanced Computing and Microelectronics Unit, Indian Statistical Institute & Department of Mathematics and Computer Science, Mizoram University (MZU), Aizawl during March 6-11, 2017.
2. Attended "ISI-MZU SCHOOL ON SOFT COMPUTING TECHNIQUES: THEORY AND APPLICATIONS" organized jointly by Machine Intelligence Unit, Indian Statistical Institute, Kolkata, and Department of Mathematics & Computer Science, Mizoram University, Aizawl during March 20-24, 2017.
3. Attended "National Workshop on Research Methodology (Qualitative & Quantitative With Advance Statistics)" organized by the School of Social Sciences, Mizoram University in collaboration with the Human Resource Development Centre (HRDC), Mizoram University during 7^{th} – 11^{th} August, 2018.
4. Attended the Instructional School for Teachers on "Mathematical Modelling in Continuum Mechanics and Ecology" organized by National Centre for Mathematics, a joint initiative of IIT Bombay & TIFR Mumbai, held at Mizoram University during June 03 – 15, 2019.
5. Attended the National Workshop on 'Ethics in Research and Preventing Plagiarism (ERPP 2019)' on 3^{rd} October 2019).

LIST OF PUBLICATIONS

1. Hussain, J., & Rosangliana, D. (2018). Analysis of Higher Education Enrolment in Mizoram. *Mizoram Educational Journal*, *4(1)*:13-24.
2. Hussain, J., Rosangliana, D., & Vanlalruata. (2020). Gross Enrolment Ratio Prediction using Artificial Neural Network. *International Journal of Recent Technology and Engineering (IJRTE)*, *8(5)*:5006-5011.
3. Hussain, J., Rosangliana, D., & Vanlalruata. Student Gross Enrolment Ratio Forecasting: A Comparative Study Using Statistical Method and Machine Learning. *International Journal of Information and Education Technology. Vol. 13, No. 3, March 2023.*

PARTICULARS OF THE CANDIDATES

Name of the Candidate:	David Rosangliana
Degree:	Doctor of Philosophy
Department:	Mathematics and Computer Science
Title of the Thesis:	Mathematical and Statistical Analysis of Higher Education in Mizoram
Date of Admission:	1.2.2012

Approval of the Research Proposal

BOS	06.05.2015
School Board	11.05.2015
MZU Registration No.	62 of 2011
Ph.D. Registration No.	MZU/Ph.D./742 of 11.5.2015
Date of Extension (if any)	For a period of 2 years up to 10.05.2022 vide AC:4(3) dated 23.06.2020
	Up to 30th June vide No. 17-16/MZU(Acad)/20/dated 12th August 2021
	Up to 31st December vide No. 17-1/MZU(Acad/20/36 dated 26th May 2022

Date : 06.06.2023 Head

Place : Aizawl Department of Mathematics and Computer Science

REFERENCES

Adeyemi, K., &Akpotu, N. (2004).Gender analysis of student enrolment in Nigerian universities. *Higher Education*, *48*(3), 361-378.

Agasisti, T., & Bertoletti, A. (2020). Higher education and economic growth: A longitudinal study of European regions 2000–2017. *Socio-Economic Planning Sciences*, 100940.

Agrawal, V. (2014). Higher Education in India: Challenges and Opportunities. *International Journal of Scientific Research*, *3*(1).

Alam, G. M., Mishra, P. K., & Shahjamal, M. M. (2014). Quality assurance strategies for affiliated institutions of HE: A case study of the affiliates under National University of Bangladesh. *Higher Education*, *68*(2), 285-301.

All India Survey on Higher Education (2010-11). New Delhi: Ministry of Education. Retrieved from https://www.education.gov.in/sites/upload_files/mhrd/files/statistics-new/AISHE2010-11.pdf

All India Survey on Higher Education (2011-12). New Delhi: Ministry of Education. Retrieved from https://www.education.gov.in/sites/upload_files/mhrd/files/statistics-new/AISHE2011-12.pdf

All India Survey on Higher Education (2012-13). New Delhi: Ministry of Education. Retrieved from https://aishe.gov.in/aishe/viewDocument.action?documentId=194

All India Survey on Higher Education (2013-14). New Delhi: Ministry of Education. Retrieved from https://www.education.gov.in/sites/uploadfiles/mhrd/files/statistics-new/AISHE2013-14.pdf

All India Survey on Higher Education (2014-15). New Delhi: Ministry of Education. Retrieved from https://aishe.gov.in/aishe/viewDocument.action?documentId=206

All India Survey on Higher Education (2015-16). New Delhi: Ministry of Education. Retrieved from https://www.education.gov.in/sites/upload_files/mhrd/files/statistics-new/AISHE2015-16.pdf

All India Survey on Higher Education (2016-17). New Delhi: Ministry of Education. Retrieved from https://aishe.gov.in/aishe/viewDocument.action?documentId=239

All India Survey on Higher Education (2017-18). New Delhi: Ministry of Education. Retrieved from https://aishe.gov.in/aishe/viewDocument.action?documentId=245

All India Survey on Higher Education (2018-19). New Delhi: Ministry of Education. Retrieved from https://aishe.gov.in/aishe/viewDocument.action?documentId=262

All India Survey on Higher Education (2019-20). New Delhi: Ministry of Education. Retrieved from https://aishe.gov.in/aishe/viewDocument.action?documentId=277

Amirtham, N. S., & Kundupuzhakkal, S. (2013). Gender issues and dropout rates in India: Major barrier in providing education for all. *Education Confab*, *2*(4).

Barbhuiya, A. H. (2014). *Cost and Finance of Higher Education in Mizoram*. New Delhi: Concept.

Bastedo, M. N., & Jaquette, O. (2011). Running in place: Low-income students and the dynamics of higher education stratification. *Educational Evaluation and Policy Analysis*, *33*(3), 318-339.

Berdahl, R. O., Moodie, G. C., & Spitzberg, I. J. (Eds.). (1991). *Quality and access in higher education: Comparing Britain and the United States*. Open University Press.

Bhagat, R. (2013). Growth of Higher Education in J & K and its Affects. *International Journal of Educational Planning & Administration*, *3*(1), 61-67.

Bhutia, Yodida (2005). A study of higher education in the state of Sikkim. Ph.D Thesis. Retrieved from http://hdl.handle.net/10603/60572

Bote-Lorenzo, M. L., & Gómez-Sánchez, E. (2017, March). Predicting the decrease of engagement indicators in a MOOC. In *Proceedings of the Seventh international learning analytics & knowledge conference* (pp. 143-147).

Bousnguar, H., Najdi, L., & Battou, A. (2022). Forecasting approaches in a higher education setting. *Education and Information Technologies*, *27*(2), 1993-2011.

Bradley, C. L., & Renzulli, L. A. (2011). The complexity of non-completion: Being pushed or pulled to drop out of high school. *Social Forces*, *90*(2), 521-545.

Chea, P. (2019). Does higher education expansion in Cambodia make access to education more equal?. *International Journal of Educational Development*, *70*, 102075.

Cheng, X., Khomtchouk, B., Matloff, N., & Mohanty, P. (2018). Polynomial regression as an alternative to neural nets. *arXiv preprint arXiv:1806.06850*.

Chuaungo, L. (2015). Access to higher education in India: An overview, Mizoram Education Journal.

Chugh, S. (2011). Dropout in secondary education: A study of children living in slums of Delhi.

Coleman J.S., (1966). *Equality of Edu. Oppor. Washington:* U.S. Office of Education.

Dadhich, M., Pahwa, M. S., Jain, V., & Doshi, R. (2021). Predictive models for stock market index using stochastic time series ARIMA modeling in emerging economy. In *Advances in mechanical engineering* (pp. 281-290). Springer, Singapore.

Dalipi, F., Imran, A. S., & Kastrati, Z. (2018, April). MOOC dropout prediction using machine learning techniques: Review and research challenges. In *2018 IEEE global engineering education conference (EDUCON)* (pp. 1007-1014). IEEE.

de Araújo Morais, L. R., & da Silva Gomes, G. S. (2022). Forecasting daily Covid-19 cases in the world with a hybrid ARIMA and neural network model. *Applied Soft Computing*, *126*, 109315.

Definition of gross enrolment ratio. Retrieved from http://uis.unesco.org/en/glossary-term//gross-enrolment-ratio.

DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (2006). An integrated model of application, admission, enrollment, and financial aid. *The Journal of Higher Education*, *77*(3), 381-429.

Directorate of Economics & Statistics, (2022). Mizoram Statistical Handbook Mizoram 2020. Retrieved from https://des.mizoram.gov.in/uploads/attachments/c17bbd8f0cb516b82ae2583ec5701da0/pages-223-2020.pdf

Directorate of Economics & Statistics, Government of Mizoram (2022).Statistical Abstract of Mizoram: 2021. Retrieved from https://des.mizoram.gov.in/uploads/attachments/2022/10/5ddf63c28bb0223c22fe291a605c5668/mizoram-statistical-abstract-2021-22082022-edited.pdf

Dubey, A. (2008). Determinants of post-higher secondary enrolment in India. *Higher Education in India*, 139.

Dziallas, M., & Blind, K. (2019). Innovation indicators throughout the innovation process: An extensive literature analysis. *Technovation*, *80*, 3-29.

Epstein, E. H. (1992). Social paradoxes of American education. *Oxford Review of Education*, *18*(3), 201-212.

Feinberg, W., & Soltis, J. F. (1998). School and Society Thinking.

Gaikwad, B. R., & Solunke, R. S. (2013). Growth of higher education in India. *International Research Journal of Social Sciences*, *2*(8), 58-60.

Global Education Monitoring Report (2015), The Education for All Development Index.

Global Education Monitoring Report (2015), The Education for All Development Index.

Gouda, S., & Sekher, T. V. (2014). Factors leading to school dropouts in India: An analysis of national family health survey-3 data. *IOSR Journal of Research & Method in Education*, *4*(6), 75-83.

Goyal, E., Purohit, S., & Bhagat, M. (2011). Study of satisfaction and usability of the Internet on studentâ s performance. *International Journal of Education and Development using ICT*, *7*(1).

Grade Break up of Institutions accredited by NAAC. Retrieved from http://naac.gov.in/images/docs/STATISTICS/Grade_break_up_-as_on_21062022.xls

Gupta, D., & Gupta, N. (2012). Higher education in India: structure, statistics and challenges. *Journal of education and Practice*, *3*(2).

Hart, G. L., Mueller, T., Toher, C., & Curtarolo, S. (2021). Machine learning for alloys. *Nature Reviews Materials*, *6*(8), 730-755.

Hashim, A. S., Awadh, W. A., & Hamoud, A. K. (2020, November). Student performance prediction model based on supervised machine learning algorithms. In *IOP Conference Series: Materials Science and Engineering* (Vol. 928, No. 3, p. 032019). IOP Publishing.

Higher Education in India: Vision 2030. (2013). FICCI Higher Education Summit 2013. Ministry of HRD & Planning Commission, Government of India.

Hone, K. S., & El Said, G. R. (2016). Exploring the factors affecting MOOC retention: A survey study. *Computers & Education*, *98*, 157-168.

Hristova, M., & Zhelezarov, I. (2006). Modeling of the criteria for measurement and assessing the quality of university education. *Facta universitatis-series: Electronics and Energetics*, *19*(3), 393-404.

Huang, J. C., Ko, K. M., Shu, M. H., & Hsu, B. M. (2020). Application and comparison of several machine learning algorithms and their integration models in regression problems. *Neural Computing and Applications*, *32*(10), 5461-5469.

Jakhar, D., & Kaur, I. (2020). Artificial intelligence, machine learning and deep learning: definitions and differences. *Clinical and experimental dermatology*, *45*(1), 131-132.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, *31*(3), 685-695.

Jeelani, S. (2012). Challenges and opportunities in higher education in 12th five year plan. University News, 50(21), 4.

Jha, S., Yang, E., Almagrabi, A. O., Bashir, A. K., & Joshi, G. P. (2021). Comparative analysis of time series model and machine testing systems for crime forecasting. *Neural Computing and Applications*, *33*(17), 10621-10636.

Johnes, J., & Taylor, J. (1989). Undergraduate non-completion rates: differences between UK universities. *Higher Education*, *18*(2), 209-225.

Jordan, W. J., Lara, J., & McPartland, J. M. (1996). Exploring the causes of early dropout among race-ethnic and gender groups. *Youth & Society*, *28*(1), 62-94.

Kale, R.K. (2006). Higher education and development of the nation. University News, 44 (33), 1.

Kareena, B., & Manoj, K. D. (2011). A demand of value based higher education system in India: A comparative study. *Journal of public Administration and policy Research*, *3*(5), 156-173.

Kaur, J., & Tagore, R. (2014). Journey of Higher Education in India "From Pre Globalisation to Globalised Era". *India: International Journal of Educational Planning and Administration*, *4*(1).

Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., ... & Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA, and ANCOVA analyses. *Review of educational research*, *68*(3), 350-386.

Kizilcec, René F., and Sherif Halawa. "Attrition and achievement gaps in online learning." In *Proceedings of the second (2015) ACM conference on learning@ scale*, pp. 57-66. 2015.

Konwar, N., & Chakraborty, S. (2013). Higher education scenario of the North-Eastern India. *Indian Journal of Research*, *2*(3).

Kumar, D.S. (2012). Recent reforms in education in India: Achievements and unfinished tasks. International Journal of Social Science & Interdisciplinary Research, 1(8), 82.

Kumar, S. S., Kumar, A., Agarwal, S., Syafrullah, M., & Adiyarta, K. (2022, October). Forecasting indoor temperature for smart buildings with ARIMA, SARIMAX, and LSTM: A fusion approach. In *2022 9th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)* (pp. 186-192). IEEE.

Lan, W., & Lanthier, R. (2003). Changes in students' academic performance and perceptions of school and self before dropping out of schools. *Journal of Education for Students Placed at Risk*, *8*(3), 309-332.

Lellep, M., Prexl, J., Linkmann, M., & Eckhardt, B. (2020). Using machine learning to predict extreme events in the Hénon map. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *30*(1), 013113.

Lillicrap, T. P., Cownden, D., Tweed, D. B., & Akerman, C. J. (2016). Random synaptic feedback weights support error backpropagation for deep learning. *Nature communications*, *7*(1), 1-10.

Liu, S., Lu, M., Li, H., & Zuo, Y. (2019). Prediction of gene expression patterns with generalized linear regression model. *Frontiers in Genetics*, *10*, 120.

Lu, M., & Zhang, X. (2019). Towards an intelligent country: China's higher education expansion and rural children's senior high school participation. *Economic Systems*, *43*(2), 100694.

Majumder, A., & Mitra, C. (2020). Dropout behaviour of children: The case of West Bengal. *Indian Journal of Human Development*, *14*(2), 275-289.

Marphatia, A. A., Reid, A. M., & Yajnik, C. S. (2019). Developmental origins of secondary school dropout in rural India and its differential consequences by sex: A biosocial life-course analysis. *International Journal of Educational Development*, *66*, 8-23.

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

McNeal Jr, R. B. (1997). Are students being pulled out of high school? The effect of adolescent employment on dropping out. *Sociology of Education*, 206-220.

Meier, V., & Schiopu, I. (2020). Enrollment expansion and quality differentiation across higher education systems. *Economic Modelling*, *90*, 43-53.

Misra, S. N., & Bal, M. (2014). Excellence through equity: Challenges before higher education in India. *International Journal of Educational Planning & Administration*, *4*(1), 67-77.

Mizoram Statistical Handbook (2020). Directorate of Economics & Statistics, Mizoram : Aizawl. Retrieved from https://des.mizoram.gov.in/uploads/attachments/c17bbd8f0cb516b82ae2583ec5701da0/pages-223-2020.pdf.

Mugo, M. G. (1999). African culture in education for sustainable development. *African renaissance*, 210-232.

Mussida, C., Sciulli, D., & Signorelli, M. (2019). Secondary school dropout and work outcomes in ten developing countries. *Journal of Policy Modeling*, *41*(4), 547-567.

Neelaveni, C., & Manimaran, S. (2014). A Statistical Study on Higher Educational Institutions in India. *International Journal of Educational Administration and Policy Studies*, *6*(9), 190-195.

NSS Report No.585: Household Social Consumption on Education in India, Ministry of Statistics and Programme Implementation, Government of India, retrieved from http://164.100.161.63/sites/default/files/publication_reports/Report_585_75th_round_Education_final_1507_0.pdf

Perna, L. W., Orosz, K., Gopaul, B., Jumakulov, Z., Ashirbekov, A., & Kishkentayeva, M. (2014). Promoting human capital development: A typology of international scholarship programs in higher education. *Educational Researcher*, *43*(2), 63-73.

Prakash, R., Beattie, T., Javalkar, P., Bhattacharjee, P., Ramanaik, S., Thalinja, R., ... & Isac, S. (2017). Correlates of school dropout and absenteeism among adolescent girls from marginalized community in north Karnataka, south India. *Journal of adolescence*, *61*, 64-76.

Prakash, V. (2007). Trends in growth and financing of higher education in India. *Economic and Political Weekly*, 3249-3258.

Puthe, M. (2008). E-learning concepts and literature review. *E-learning issues in Malaysia higher education*, 1-22.

Rahim, R., Ibrahim, H., Kasim, M. M., & Adnan, F. A. (2013). Projection model of postgraduate student flow. *Appl. Math*, *7*(2L), 383-387.

Rahmati, O., Choubin, B., Fathabadi, A., Coulon, F., Soltani, E., Shahabi, H., ... & Bui, D. T. (2019). Predicting uncertainty of machine learning models for modelling nitrate pollution of groundwater using quantile regression and UNEEC methods. *Science of the Total Environment*, *688*, 855-866.

Ramesh, G. (2013). Indian Higher Education and the Challenges of Sustainability: An Analytical Note. *International Journal of Social Science & Interdisciplinary Research*, *2*(9), 28-37.

Rao V.V., Lakshmi V.V., Krishna V.V. and Rao B.B., (2004). *Education For All,* Sonali Publications, New Delhi, p. 243-244.

Ray, S. (2019, February). A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)* (pp. 35-39). IEEE.

Ray, S., Das, S. S., Mishra, P., & Al Khatib, A. M. G. (2021). Time series SARIMA modelling and forecasting of monthly rainfall and temperature in the South Asian countries. *Earth Systems and Environment*, *5*(3), 531-546.

Reale, E. (2014). Challenges in higher education research: The use of quantitative tools in comparative analyses. *Higher Education*, *67*(4), 409-422.

Riggert, S. C., Boyle, M., Petrosko, J. M., Ash, D., & Rude-Parkins, C. (2006). Student employment and higher education: Empiricism and contradiction. *Review of educational research*, *76*(1), 63-92.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), e1355.

Rowan - Kenyon, Heather T. (2007), “Predictors of Delayed College Enrolment and the Impact of Socioeconomic Status”, The Journal of Higher Education, Vol. 78, No.2, March-April, pp. 188-214.

Rupon, B. (2012). School dropout across Indian states and UTs: an econometric study. *International Research Journal of Social Sciences*, *1*(4), 28-35.

Said, E. H. (2011). ICT use by journalism professors in Colombia. *Australasian Journal of Educational Technology*, *27*(2).

Sanusi, A., & Oyama, T. (2008). Statistical data analysis for investigating Japanese government subsidy policy for private universities. *Higher Education*, *55*(4), 407-423.

Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, *20*(1), 3-29.

Schwarz, S., & Westerheijden, D. F. (Eds.). (2007). *Accreditation and evaluation in the European higher education area* (Vol. 5). Springer Science & Business Media.

Sharma, S., & Levinson, D. (2019). Travel cost and dropout from secondary schools in Nepal. *Transportation research part A: policy and practice*, *130*, 385-397.

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2018, December). A comparison of ARIMA and LSTM in forecasting time series. In *2018 17th IEEE international conference on machine learning and applications (ICMLA)* (pp. 1394-1401). IEEE.

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). A comparative analysis of forecasting financial time series using arima, lstm, and bilstm. *arXiv preprint arXiv:1911.09512*.

Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-Means Clustering Algorithm. IEEE Access 8, 80716–80727.

Singh, K.P., & Ahmad, S. (2011). Higher education in India: Major concerns, University News. 49 (29), 1.

Singh, T. S. (2013). Problems And Prospects Of Higher Education In Manipur (No. 2013-9-3).

Single year age returns by residence and sex. Retrieved from http://www.censusindia.gov.in/2011census/Age_level-Data/India/Age_data.xls.

Sinha, S., & Srivastava, R. S. (2008). Inclusiveness and access of social groups to higher education. *Higher Education in India*, 111.

Stake, J. E., & Hoffmann, F. L. (2001). Changes in student social attitudes, activism, and personal confidence in higher education: The role of women's studies. *American educational research journal*, *38*(2), 411-436.

State Profile (2005-06). New Delhi: Ministry of Education. Retrieved from https://www.education.gov.in/sites/upload_files/mhrd/files/statistics-new/StateProfile-2005-06.pdf

State wise number of Universities and Colleges accredited by NAAC. Retrieved from http://naac.gov.in/images/docs/STATISTICS/State_wise_number_of_Universities_and_Colleges_accredited_by_NAAC_-_as_on_21062022.xls

State wise number of Universities and Colleges accredited by NAAC. Retrieved from http://naac.gov.in/images/docs/STATISTICS/State_wise_number_of_Universities_ and_Colleges_accredited_by_NAAC_-_as_on_21062022.xls

State wise number of Universities and Colleges accredited by NAAC . Retrieved from http://naac.gov.in/images/docs/STATISTICS/State_wise_number_of_Universities_ and_Colleges_accredited_by_NAAC_-_as_on_21062022.xls

Statistics of Higher and Technical Education (2006 - 2007). Ministry of Education. Retrieved from https://www.education.gov.in/sites/upload_files/mhrd/files/statistics-new/Abstract2006-07.pdf

Statistics of Higher and Technical Education (2007 - 2008). Ministry of Education. Retrieved from https://www.education.gov.in/sites/upload_files/mhrd/files/statistics-new/Stat-HTE-200708_0.pdf

Statistics of Higher and Technical Education (2008 - 2009). Ministry of Education. Retrieved from https://www.education.gov.in/sites/uploadfiles/mhrd/files/statistics-new/StatHTE_2008-09.pdf

Statistics of Higher and Technical Education (2009 - 2010). Ministry of Education. Retrieved from https://www.education.gov.in/sites/upload_files/mhrd/files/statistics-new/Abstract2009-10.pdf

Statistics of School Education (2006-07). New Delhi: Ministry of Education. Retrieved from https://www.education.gov.in/sites/upload_files/mhrd/files/statistics-new/SES-School-2006-07.pdf

Statistics of School Education (2007-08). New Delhi: Ministry of Education. Retrieved from https://www.education.gov.in/sites/upload_files/mhrd/files/statistics-new/SES-School-2007-08.pdf

Statistics of School Education (2008-09). New Delhi: Ministry of Education. Retrieved from https://www.education.gov.in/sites/upload_files/mhrd/files/statistics-new/SES-School_2008-09.pdf

Statistics of School Education (2009-10). New Delhi: Ministry of Education. Retrieved from https://www.education.gov.in/sites/upload_files/mhrd/files/statistics-new/SES-School_2009-10.pdf

Statistics of School Education (2010-11). New Delhi: Ministry of Education. Retrieved from https://www.education.gov.in/sites/upload_files/mhrd/files/statistics-new/SES-school-2011-12.pdf

Stearns, E., & Glennie, E. J. (2006). When and why dropouts leave high school. *Youth & Society*, *38*(1), 29-57.

Strauss, L. M., & Borenstein, D. (2015). A system dynamics model for long-term planning of the undergraduate education in Brazil. *Higher Education*, *69*(3), 375-397.

Thangaraj, C. (2016). Indian higher education: Contributions to human capital, competitiveness and social progression. University News, 54 (51), 27.

Thorat, S. (2010). Emerging Issues in Higher Education- Approach and Strategy. In Khurana, S.M.P., & Singhal, P.K. (E.d) Higher Education Quality and Management, New Delhi:Gyan Publishing House, 385.

Türkkahraman, M. (2012). The role of education in the societal development. *Journal of educational and instructional studies in the world*, *2*(4), 38-41.

UDISE Plus data on enrolment and dropout (2012-2022). Ministry of Education. Retrieved from https://dashboard.udiseplus.gov.in/#/reportDashboard/sReport

Unit level data of NSSO round 75 Schedule 25.2: Household Social Consumption: Education, retrieved from https://mospi.gov.in/web/mospi/download-tables-data/-/reports/view/templateTwo/16203?q=TBDCAT

Uralov, O. S. (2020). Internationalization of higher education in Uzbekistan. *Social Sciences & Humanities Open*, *2*(1), 100015.

Van Vught, F. A., & Westerheijden, D. F. (1994). Towards a general model of quality assessment in higher education. *Higher education*, *28*(3), 355-371.

Vyas, U., & Basu, S. (2009). Higher education in India: Reality check. University News, 47(21).

Weybright, E. H., Caldwell, L. L., Xie, H., Wegner, L., & Smith, E. A. (2017). Predicting secondary school dropout among South African adolescents: A survival analysis approach. *South African journal of education*, *37*(2), 1-11.

Wild, S., & Heuling, L. S. (2020). Student dropout and retention: An event history analysis among students in cooperative higher education. *International Journal of Educational Research*, *104*, 101687.

Yadav, M.S. (1985). Higher education in India: Survival and revival. Indian Educational Review. 20(3), 1.

Yang, S., Chen, H. C., Chen, W. C., & Yang, C. H. (2020). Student enrollment and teacher statistics forecasting based on time-series analysis. *Computational intelligence and neuroscience*, *2020*.

Ye, C., Yang, J., & Ding, H. (2020). Bagging for Gaussian mixture regression in robot learning from demonstration. *Journal of Intelligent Manufacturing*, 1-13.

Zhang, F., Zhao, Y., Zhang, S., Wu, W., & Tan, C. (2021, August). Spacecraft Equipment Health Condition Monitoring Based on Augmented Dickey-Fuller Test and Gaussian Mixture Model. In *2021 IEEE International Conference on Mechatronics and Automation (ICMA)* (pp. 1379-1384). IEEE.

Zmas, A. (2015). Financial crisis and higher education policies in Greece: Between intra-and supranational pressures. *Higher Education*, *69*(3), 495-508.

Zohmingliani L., (2012). *Status of Science education in Mizoram: A critical study,* Phd Thesis.