

**KOMBINASI RANGKAIAN NEURAL BUATAN  
DENGAN MODEL LINEAR: APLIKASINYA  
DALAM SAINS KESIHATAN**

**FARAH MUNA BINTI MOHAMAD GHAZALI**

**UNIVERSITI SAINS MALAYSIA**

**2024**

**KOMBINASI RANGKAIAN NEURAL BUATAN  
DENGAN MODEL LINEAR: APLIKASINYA  
DALAM SAINS KESIHATAN**

oleh

**FARAH MUNA BINTI MOHAMAD GHAZALI**

**Tesis ini diserahkan untuk  
memenuhi keperluan bagi  
Ijazah Doktor Falsafah**

**Mei 2024**

## PENGHARGAAN

السَّلَامُ عَلَيْكُمْ وَرَحْمَةُ اللَّهِ وَبَرَكَاتُهُ

Terlebih dahulu saya ingin mengambil kesempatan untuk memanjatkan rasa kesyukuran ke hadrat Allah Yang Maha Kuasa kerana dengan limpah dan izin-Nya, telah memberi petunjuk, ilham, kesihatan yang baik, masa, serta kekuatan kepada saya bagi menyempurnakan tesis saya ini dengan jayanya. Selawat dan Salam ke atas junjungan besar Nabi Muhammad SAW. Setinggi-tinggi penghargaan dan ucapan terima kasih khas saya tujukan kepada penyelia saya, Prof. Madya Ts. Dr. Wan Muhamad Amir bin W Ahmad, yang memberikan bimbingan dan galakan serta mendorong semangat saya untuk menulis tesis dengan penuh dedikasi. Saya bangga memiliki penyelia sebaik beliau yang tidak lekang berkongsi ilmu, terutama dalam bidang penerbitan buku dan jurnal. Saya juga ingin merakamkan jutaan terima kasih kepada Pusat Pengajian Sains Pergigian dan Universiti Sains Malaysia atas penajaan Skim Siswazah Pembantu sepanjang pengajian, serta terima kasih kepada Kementerian Pengajian Tinggi Malaysia atas bantuan dana Skim Geran Penyelidikan Asas (Projek Kod: FRGS/1/2022/STG06/USM/02/10). Penghargaan juga disampaikan kepada ibu tersayang, Pn. Hajah Hazizah binti Dollah, suami yang tercinta, Wan Mohd Fakhzan bin Wan Deraman, adik-beradik, keluarga mentua, dan semua ahli keluarga yang memberikan sokongan, dorongan, dan galakan sepanjang tempoh pengajian sehingga peringkat Doktor Falsafah. Akhir sekali, ucapan ribuan terima kasih kepada sahabat-sahabat seperjuangan di USM, khususnya Dr. Muhammad Azeem Yaqoob, dan Mohamad Nasarudin bin Adnan, serta Dr. Rabiatul Adawiyah binti Abdul Rohim dari UniSZA, dan kepada semua yang turut membantu dan memberi pandangan untuk penambahbaikan tesis ini. Semoga ilmu yang diperoleh memberi manfaat dan berkat kepada semua. Jutaan terima kasih atas segala jasa baik yang telah diberikan.

## JADUAL KANDUNGAN

<b>PENGHARGAAN</b> .....	<b>ii</b>
<b>JADUAL KANDUNGAN</b> .....	<b>iii</b>
<b>SENARAI JADUAL</b> .....	<b>vii</b>
<b>SENARAI RAJAH</b> .....	<b>viii</b>
<b>SENARAI SINGKATAN</b> .....	<b>x</b>
<b>ABSTRAK</b> .....	<b>xi</b>
<b>ABSTRACT</b> .....	<b>xii</b>
<b>BAB 1 PENDAHULUAN</b> .....	<b>1</b>
1.1 Pengenalan bab.....	1
1.2 Pengenalan kepada pemodelan rangkaian neural dalam sains kesihatan.....	1
1.3 Keperluan pemodelan model rangkaian neural berintegrasi .....	3
1.4 Latar belakang dan perkembangan penyelidikan .....	4
1.4.1 Keperluan pemodelan secara model regresi linear berintegrasi dengan model rangkaian neural dalam bidang sains kesihatan .....	8
1.5 Pernyataan masalah.....	10
1.6 Rasional penyelidikan .....	12
1.7 Hipotesis penyelidikan .....	13
1.8 Objektif penyelidikan umum.....	13
1.8.1 Objektif penyelidikan khusus .....	13
1.9 Persoalan penyelidikan.....	14
1.10 Skop penyelidikan kajian .....	14
1.11 Kerangka penyelidikan keseluruhan .....	15
1.12 Sumbangan kajian .....	16
1.13 Keterbatasan kajian penyelidikan .....	17
1.14 Organisasi tesis.....	18

<b>BAB 2</b>	<b>SOROTAN KESUSASTERAAN .....</b>	<b>21</b>
2.1	Pengenalan bab.....	21
2.2	Bahagian I: Sejarah dan perkembangan model rangkaian neural .....	21
2.2.1	Pendekatan rangkaian neural dalam bidang sains kesihatan .....	29
2.2.2	Tinjauan aplikasi model rangkaian neural.....	32
2.2.3	Kelebihan rangkaian neural.....	34
2.3	Bahagian II: Sejarah dan perkembangan kaedah model regresi linear .....	35
2.3.1	Tinjauan terhadap aplikasi regresi linear.....	37
2.3.2	Kelebihan regresi linear.....	40
2.4	Bahagian III: Kaedah butstrap .....	41
2.5	Bahagian IV: Antara muka pengguna grafik ( <i>GUI</i> ).....	42
2.6	Bahagian V: Diabetes mellitus.....	44
2.7	Bahagian VI: Aplikasi kaedah statistik dalam sains perubatan dan kesihatan.....	46
2.7.1	Kajian literatur berkaitan dengan pemodelan aras kreatinin .....	47
2.7.2	Kajian literatur berkaitan dengan indeks jisim badan ( <i>BMI</i> ).....	49
2.7.3	Kajian literatur berkaitan dengan pemodelan aras <i>LDL</i> .....	51
2.8	Bahagian VI: Tinjauan terhadap metodologi dan batasan kajian lepas .....	51
2.9	Perumusan .....	53
<b>BAB 3</b>	<b>METODOLOGI .....</b>	<b>55</b>
3.1	Pengenalan bab.....	55
3.2	Reka bentuk kajian .....	55
3.2.1	Statistik komputeran .....	56
3.3	Sampel dan populasi kajian.....	57
3.4	Lokasi penyelidikan .....	58
3.5	Jangka masa kajian penyelidikan .....	58
3.6	Penentuan saiz sampel kajian.....	58
3.7	Teknik persampelan .....	58

3.8	Kelulusan etika.....	58
3.9	Perisian yang digunakan dalam kajian penyelidikan .....	59
3.10	Penentuan pemboleh ubah dalam kajian kes.....	59
3.10.1	Kajian kes I: Pemodelan aras kreatinin dalam kalangan pesakit dislipidemia dengan diabetes mellitus jenis 2 .....	59
3.10.2	Kajian kes II: Pemodelan indeks jisim badan ( <i>BMI</i> ) dalam kalangan pesakit diabetes mellitus jenis 2.....	60
3.10.3	Kajian kes III: Pemodelan terhadap bacaan aras lipoprotein berketumpatan rendah ( <i>LDL</i> ) dalam kalangan pesakit diabetes mellitus jenis 2.....	61
3.11	Pengenalan algoritma menggunakan bahasa pengaturcaraan .....	62
3.11.1	Bahasa pengaturcaraan <i>R</i> .....	63
3.11.2	Tatacara perencanaan input data ke dalam perisian <i>RStudio</i> dan kelebihan <i>R</i> .....	64
3.11.3	Bahasa pengaturcaraan <i>Java</i> .....	67
3.11.4	Tatacara perencanaan input data dalam perisian <i>Eclipse</i> dan kelebihan <i>Java</i> .....	69
3.12	Model biostatistik komputasi .....	76
3.13	Komponen-Komponen utama bagi pembinaan metodologi dengan menggunakan bahasa pengaturcaraan <i>R</i> dan <i>Java</i> .....	76
3.13.1	Bahagian I: Pengenalan kepada kaedah butstrap.....	76
3.13.2	Bahagian II: Pengenalan kepada model rangkaian neural hadapan suapan berbilang lapisan ( <i>MLFFNN</i> ).....	77
3.13.3	Bahagian III: Pengenalan kepada pemodelan linear .....	79
3.13.3(a)	Teori yang mendasari kaedah regresi linear .....	79
3.13.4	Bahagian IV: Pengenalan kepada pembangunan <i>GUI</i> .....	80
3.14	Prosedur kaedah gabungan model <i>LiReNN</i> .....	81
3.14.1	Tatacara pelaksanaan algoritma model <i>LiReNN</i> .....	81
3.14.2	Carta alir pembentukan model <i>LiReNN</i> .....	92
3.14.3	Rumusan.....	92
3.15	Kesimpulan .....	93

<b>BAB 4</b>	<b>KEPUTUSAN DAN PERBINCANGAN.....</b>	<b>95</b>
4.1	Pengenalan bab.....	95
4.2	Penerangan kajian kes I.....	96
4.2.1	Analisis dan dapatan kajian kes I.....	96
4.2.2	Perumusan kajian kes I.....	103
4.3	Penerangan kajian kes II .....	104
4.3.1	Analisis dan dapatan kajian kes II.....	104
4.3.2	Perumusan kajian kes II .....	109
4.4	Penerangan kajian kes III .....	110
4.4.1	Analisis dan dapatan kajian kes III .....	111
4.4.2	Perumusan kajian kes III.....	116
4.5	Kesimpulan .....	119
<b>BAB 5</b>	<b>PERBINCANGAN DAN KESIMPULAN .....</b>	<b>120</b>
5.1	Pengenalan bab.....	120
5.2	Kebaharuan dalam kaedah <i>LiReNN</i> .....	120
5.3	Ringkasan kajian .....	123
5.4	Kesimpulan .....	125
<b>BAB 6</b>	<b>BATASAN DAN MANFAAT KAJIAN .....</b>	<b>127</b>
6.1	Pengenalan bab.....	127
6.2	Batasan kajian .....	127
6.3	Manfaat hasil daripada kajian .....	128
6.4	Cadangan kajian pada masa hadapan.....	131
	<b>SENARAI RUJUKAN .....</b>	<b>133</b>
	LAMPIRAN	
	SENARAI PENERBITAN DAN PENYELIDIKAN	
	MAKLUMAT PERIBADI	

## SENARAI JADUAL

### Muka surat

Jadual 2.1: Aplikasi rangkaian neural dalam bidang sains kesihatan.....	33
Jadual 2.2: Ringkasan aplikasi regresi linear mudah.....	38
Jadual 2.3: Ringkasan aplikasi regresi linear berganda .....	38
Jadual 4.1: Keputusan regresi linear berganda diperoleh dengan menggabungkan set data latihan dan ujian menggunakan kaedah butstrap .....	99
Jadual 4.2: Nilai “Sebenar” dan “Ramalan” diperoleh melalui metodologi yang dicadangkan.....	99
Jadual 4.3: Keputusan regresi linear berganda diperoleh dengan menggabungkan set data latihan dan ujian menggunakan kaedah butstrap .....	107
Jadual 4.4: Nilai “Sebenar” dan “Ramalan” diperoleh melalui metodologi yang dicadangkan.....	107
Jadual 4.5: Keputusan regresi linear berganda diperoleh dengan menggabungkan set data latihan dan ujian menggunakan kaedah butstrap .....	114
Jadual 4.6: Nilai “Sebenar” dan “Ramalan” diperoleh melalui metodologi yang dicadangkan.....	114



## SENARAI RAJAH

### Muka surat

Rajah 1.1: Kerangka kajian secara keseluruhan .....	15
Rajah 2.1: Rangkaian neuron biologi (Sumber: Gill, 2023).....	23
Rajah 2.2: Struktur rangkaian neural buatan (Sumber: Siddharth, 2021).....	24
Rajah 3.1: Model umum <i>MLFFNN</i> dengan satu lapisan tersembunyi, $N$ nod masukan, $H$ nod tersembunyi, dan satu nod keluaran .....	78
Rajah 3.2: Halaman log masuk.....	88
Rajah 3.3: Masukkan nama pengguna dan kata laluan.....	89
Rajah 3.4: Pangkalan data ( <i>database</i> ) .....	90
Rajah 3.5: Halaman pendaftaran pengguna baru .....	90
Rajah 3.6: Berjaya log masuk.....	90
Rajah 3.7: Halaman utama sistem <i>LiReNN</i> .....	91
Rajah 3.8: Carta alir tatacara pelaksanaan model <i>LiReNN</i> .....	92
Rajah 4.1: Seni bina <i>MLFFNN</i> terdiri daripada tiga nod masukan, satu lapisan tersembunyi dan satu nod keluaran .....	97
Rajah 4.2: Halaman utama <i>LIRENN</i> .....	100
Rajah 4.3: Mesej maklumat kajian kes I.....	100
Rajah 4.4: Keputusan kajian kes I .....	101
Rajah 4.5: Plot kajian kes I.....	102
Rajah 4.6: Storan untuk membuka fail <i>R</i> .....	102
Rajah 4.7: Aplikasi <i>RStudio</i> .....	103
Rajah 4.8: Fail <i>R</i> dibuka di <i>RStudio</i> .....	103
Rajah 4.9: Seni bina <i>MLFFNN</i> terdiri daripada lapan nod masukan, dua lapisan tersembunyi dan satu nod keluaran .....	106
Rajah 4.10: Mesej maklumat kajian kes II .....	108

Rajah 4.11: Keputusan kajian kes II.....	108
Rajah 4.12: Plot kajian kes II.....	109
Rajah 4.13: Seni bina <i>MLFFNN</i> terdiri daripada enam nod masukan, satu lapisan tersembunyi dan satu nod keluaran .....	113
Rajah 4.14: Mesej maklumat kajian kes III.....	115
Rajah 4.15: Keputusan kajian kes III.....	115
Rajah 4.16: Plot kajian kes III .....	116
Rajah 4.17: Mesej setelah klik butang log keluar.....	116

## SENARAI SINGKATAN

ALP	Alkaline Phosphatase
Alt	Alanine Transferase
ANN	Artificial Neural Network
BMI	Body mass index
BNN	Biological Neural Network
Chd	Coronary Heart Disease
Dys	Dyslipidemia
GUI	Graphical User Interface
HCI	Human-Computer Interaction
HDL	High-Density Lipoprotein
LDL	Low-Density Lipoprotein
LM	Linear Model
ML	Machine Learning
MLFFNN	Multi-Layer Feed-Forward Neural Network
MLR	Multiple Linear Regression
MLP	Multi-Layer Perceptron
MSE	Mean Square Error
MSE.lm	Mean Square Error Linear Model
MSE.net	Mean Squared Error Neural Network
R	Bahasa Pengaturcaraan R
LiReNN	Linear Regression Neural Network
IPS	Institut Pengajian Siswazah
RLB	Regresi Linear Berganda
PPSG	Pusat Pengajian Sains Pergigian
RNB	Rangkaian Neural Buatan
T2DM	Diabetes Mellitus Jenis 2
Tc	Total Cholesterol
Tg	Triglycerides
USM	Universiti Sains Malaysia
Wc	Waist Circumference

# KOMBINASI RANGKAIAN NEURAL BUATAN DENGAN MODEL LINEAR: APLIKASINYA DALAM SAINS KESIHATAN

## ABSTRAK

Pembangunan metodologi yang mengintegrasikan model linear dan rangkaian neural buatan dalam bidang sains kesihatan adalah suatu pendekatan yang menarik untuk meningkatkan kualiti pemodelan berserta ramalan. Kajian sebelum ini cenderung memberi penekanan kepada teknik tunggal, yang terbukti tidak mencukupi secara komprehensif. Justeru, terdapat keperluan secara statistik yang lebih holistik yang menggabungkan beberapa teknik, seperti model *LiReNN*, untuk meningkatkan ketepatan dalam pemodelan sains kesihatan. Kajian ini bertujuan untuk membangunkan model regresi linear dengan rangkaian neural yang menumpukan kepada kecekapan, tahap kompetensi, membuat peramalan, dan mengesahkan ketepatan model, serta mempermudah proses analisis data melalui antara muka pengguna grafik (*GUI*). Pendekatan gabungan ini melibatkan penggunaan butstrap, regresi linear, dan rangkaian neural hadapan suapan berbilang lapisan, disokong oleh *GUI*. Ketepatan kaedah dinilai melalui nilai *MSE.net* dan  $R^2$ , menunjukkan peningkatan yang signifikan dalam ketepatan, keberkesanan, dan kecekapan analisis data. Hasil ini membantu dalam pembangunan metodologi penyelidikan yang lebih kukuh, menyokong keputusan yang lebih tepat, dan menyediakan penggunaan *GUI* yang lebih interaktif. Pendekatan ini meningkatkan keberkesanan aplikasi pemodelan *LiReNN*, membolehkan pengguna untuk mencapai hasil optimum tanpa keperluan kepada pengetahuan mendalam dalam pengaturcaraan atau analisis data. Secara keseluruhan, pendekatan ini tidak hanya meningkatkan ketepatan dan kecekapan analisis data, tetapi juga menyediakan antara muka pengguna yang lebih mesra pengguna dan efisien melalui penggunaan *GUI*.

# COMBINATION OF ARTIFICIAL NEURAL NETWORK AND LINEAR REGRESSION: ITS APPLICATION ON HEALTH SCIENCES

## ABSTRACT

The development of a methodology that integrates linear models and artificial neural networks in the field of health sciences is an intriguing approach to improving the quality of modeling with forecasting. Previous studies have tended to emphasize single techniques, which have been proven to be insufficient comprehensively. Therefore, there is a need for a more holistic statistical approach that combines various techniques, such as the *LiReNN* model, to improve accuracy in health science modeling. This study aims to develop a linear regression model with a neural network focusing on efficiency, competency level, forecasting, and verifying model accuracy, as well as facilitating data analysis through a graphical user interface (*GUI*). This combined approach involves the use of bootstrap, linear regression, and multi-layer feedforward neural networks, supported by a *GUI*. Method accuracy is assessed through *MSE.net* and  $R^2$  values, showing a significant improvement in accuracy, effectiveness, and efficiency of data analysis. These results aid in the development of a stronger research methodology, supporting more precise decisions, and providing more interactive *GUI* usage. This approach enhances the effectiveness of *LiReNN* modeling applications, allowing users to achieve optimal results without requiring in-depth knowledge of programming or data analysis. Overall, this approach not only improves the accuracy and efficiency of data analysis but also provides a user-friendly and efficient *GUI*.

## **BAB 1**

### **PENDAHULUAN**

#### **1.1 Pengenalan bab**

Bab satu menerangkan secara terperinci keseluruhan projek penyelidikan yang telah dijalankan, bersertakan dengan beberapa subtajuk, bermula dengan perbincangan mengenai pengenalan dan keperluan pemodelan rangkaian neural dalam bidang sains kesihatan. Selepas itu, latar belakang kajian dan perkembangan penyelidikan turut diterangkan secara terperinci. Pernyataan masalah dan rasional penyelidikan dibincangkan dalam bahagian seterusnya, diikuti dengan hipotesis penyelidikan. Bahagian seterusnya, memberi penjelasan mengenai objektif penyelidikan yang merangkumi objektif umum dan juga objektif khusus. Di samping itu, bab ini juga mempersembahkan persoalan penyelidikan, skop penyelidikan, kerangka penyelidikan membentangkan gambar rajah untuk keseluruhan penyelidikan, dan sumbangan penyelidikan bagi memenuhi keperluan penyelidikan yang akan dilaksanakan. Keterbatasan kajian turut berikan dalam bahagian akhir bab ini, dan diikuti oleh organisasi tesis secara keseluruhan.

#### **1.2 Pengenalan kepada pemodelan rangkaian neural dalam sains kesihatan**

Rangkaian neural buatan, juga dikenali sebagai rangkaian neural atau singkatannya *ANN*, mewakili kerangka kerja komputasi yang berasaskan model matematik. Model-model ini berbeza daripada kaedah komputasi tradisional kerana struktur dan operasinya yang mirip dengan fungsi utama otak manusia (James, 2021). Selain itu, rangkaian neural juga dikenali sebagai sistem perhubungan, sistem teragih selari, atau sistem penyesuaian kerana ia terdiri daripada elemen pemprosesan yang berhubungan antara satu sama lain yang berfungsi secara serentak. Berbeza daripada sistem konvensional yang mempunyai kawalan berpusat, rangkaian neural

menunjukkan penyesuaian serentak semua elemen pemrosesan yang berhubungan ketika berinteraksi dengan maklumat dan peraturan penyesuaian (Sordo, 2002; Pantic *et al.*, 2023).

Pada awalnya, matlamat utama rangkaian neural buatan adalah untuk mengkaji dan memberi kesan bagaimana otak manusia berfungsi dan memproses maklumat seperti pengesanan deria, mengkategorikan idea, menghubungkan konsep, dan pembelajaran. Walau bagaimanapun, Sordo (2002) dan Akbarian *et al.* (2023) menyatakan bahawa pada masa ini terdapat banyak penekanan yang diberikan pada kemajuan rangkaian neural khususnya untuk aplikasi seperti pengecaman imej perubatan dan pengelasan corak, pemampatan data, peramalan dan pengoptimuman. Tambahan pula, rangkaian neural telah menjadi bidang yang semakin penting dalam sains kesihatan. Menurut Haykin (2007), Haykin (2009), dan Ul Rahman *et al.* (2023), rangkaian neural adalah merujuk kepada model matematik dan komputasi yang terinspirasi daripada struktur dan fungsi otak manusia.

Bidang sains kesihatan juga tidak ketinggalan dalam mengaplikasikan rangkaian neural, terutamanya untuk menganalisis dan memahami pelbagai aspek kesihatan, termasuklah diagnosis penyakit, penentuan prognosis, pengurusan data klinikal, dan peramalan kejadian kesihatan. Rangkaian neural membantu sistem supaya menjadi lebih cekap untuk mengendalikan data melalui pengecaman corak berserta dengan pemodelan kompleksiti yang tinggi (Ishak *et al.*, 2001; Pantic *et al.*, 2023; Akbarian *et al.*, 2023). Hal ini kerana, ianya berupaya mengesan hubungan dan corak yang tersembunyi dalam data kesihatan, yang mungkin sukar dikenal pasti oleh manusia. Dengan menggunakan rangkaian neural, para doktor dan penyelidik boleh menghasilkan model-model secara statistik yang sangat berguna yang berupaya

meramalkan tahap kesihatan pesakit, mengenal pasti risiko penyakit, dan juga memahami mekanisme penyakit secara lebih baik. Contoh penggunaan rangkaian neural dalam sains kesihatan termasuklah penggunaan alat pengesanan kanser berdasarkan imej perubatan, pengklasifikasian data genetik untuk diagnosis penyakit genetik, pengurusan data kesihatan elektronik, menganalisis data klinikal dan juga membuat peramalan terhadap keputusan sesuatu rawatan (Ishak *et al.*, 2001; Mukherjee *et al.*, 2023).

Justeru, pengenalan kepada rangkaian neural dalam sains kesihatan telah membuka satu ruang yang besar untuk mempertingkatkan ketepatan sesuatu diagnosis, rawatan, dan pengurusan kesihatan secara keseluruhan. Dengan menggunakan teknologi sebegini, adalah diharapkan keberkesanan sesuatu rawatan dapat ditambah baik, malah berkemungkinan besar boleh menghasilkan satu penemuan baru (Ishak *et al.*, 2001; Lorkowski *et al.*, 2021; Mukherjee *et al.*, 2023).

### **1.3 Keperluan pemodelan model rangkaian neural berintegrasi**

Model rangkaian neural memerlukan beberapa keperluan atau komponen utama bagi memastikan keberkesanan dan kejituan dalam memproses dan menganalisis data. Hal ini perlulah diselaraskan kerana keperluan pemodelan untuk model rangkaian neural berintegrasi boleh juga berbeza-beza bergantung kepada tugas tertentu dan juga jenis seni bina model yang diaplikasikan.

Terdapat beberapa keperluan umum dan juga beberapa faktor penting yang perlu dipertimbangkan semasa mereka bentuk pemodelan sedemikian. Antara keperluan utama adalah persediaan data, hal ini memastikan data yang berkualiti tinggi dan bersih digunakan untuk tujuan pemodelan. Untuk tujuan ini, set data dibahagikan kepada set



latihan, pengujian dan pengesahan. Hal ini untuk menilai prestasi model yang telah dibangun (LeCun *et al.*, 2015). Selepas itu, seni bina model dengan reka bentuk senibina khusus akan diintegrasikan kepada beberapa rangkaian neural, dan konfigurasi melalui lapisan *input* dan *output* berdasarkan data khusus dan jenis masalah yang dikaji (Goodfellow *et al.*, 2016).

#### **1.4 Latar belakang dan perkembangan penyelidikan**

Menurut Furrer (2023) penggunaan statistik kini diakui secara meluas dan diterima pakai sebagai komponen penting dalam penyelidikan dalam pelbagai domain saintifik, termasuk ekonomi, pendidikan, pengurusan perniagaan, perubatan dan biologi. Dalam bidang biologi dan perubatan, pendekatan statistik dikenali sebagai biostatistik ataupun biometrik, iaitu satu cabang statistik gunaan yang menggunakan teknik-teknik statistik untuk membuat sesuatu kesimpulan yang merangkumi analisis data, pemodelan dan juga peramalan. Analisis regresi adalah merupakan salah satu alat penyelidikan yang amat popular dalam mengkaji hubungan sebab musabab (Kuhn, 2023).

Pada masa kini, “analisis regresi” ditakrifkan sebagai kajian tentang hubungan antara satu pemboleh ubah dengan satu atau lebih pemboleh ubah peramal, dengan tujuan untuk menganggar dan juga meramalkan sesuatu pemboleh ubah bersandar berdasarkan kaedah kuasa dua terkecil (Ahmad *et al.*, 2018). Kaedah kuasa dua terkecil, juga turut dikenali sebagai kaedah penganggaran asal parameter yang telah dibangunkan oleh Legendre dan Gauss pada tahun 1805. Istilah “regresi” di perkenalkan oleh Francis Galton, yang menjalankan kajian yang mengkaji perhubungan ketinggian kanak-kanak dengan tahap ketinggian ibu bapa. Beliau mendapati bahawa ketinggian kanak-kanak cenderung untuk “*regress*” ke arah ketinggian purata ibu bapa

mereka (Krashniak & Lamm, 2021). Menurut Ahmad *et al.* (2017), Karl Pearson dalam kajiannya mendapati bahawa apabila kanak-kanak mempunyai tahap tinggi seiring dengan ketinggian purata populasi yang dikaji. Pada tahun 1875, Galton telah menggunakan kaedah biostatistik iaitu secara analisis regresi untuk menganalisis data yang berkaitan dengan perhubungan di antara pemboleh ubah bersandar dan pemboleh ubah tidak bersandar. Beliau telah mengkaji bahawa terdapat perkaitan di antara saiz kacang manis dengan saiz biji benih induk. Selain itu juga, beliau mendapati bahawa terdapat suatu corak secara garis lurus yang menaik, hal ini menunjukkan bahawa benih yang lebih besar datang daripada tumbuhan induk yang lebih besar. Hal ini menyebabkan beliau percaya bahawa garis lurus pada graf boleh membantu penyelidik untuk memahami sesuatu dengan lebih baik serta membuat kesimpulan bahawa kaedah garis regresi lurus dapat mewakili sesuatu keadaan serta ianya juga berpotensi untuk meramalkan sesuatu keadaan secara tepat. Bertitik tolak daripada kejadian itu, saintis telah melakukan banyak penyelidikan mengenai idea ini serta telah membuat penambahbaikan. Walau bagaimanapun, definisi regresi hari ini tidak jauh berbeza daripada definisi asal yang telah diperkenalkan oleh Galton, beliau berpendapat regresi adalah berfokuskan kepada penyuaian dua pemboleh ubah iaitu pemboleh ubah bersandar dan juga pemboleh ubah tidak bersandar.

Analisis regresi merupakan suatu metodologi statistik yang digunakan secara meluas dalam pelbagai bidang seperti kewangan, sains sosial, fizik, kimia, dan sains kesihatan. Ia bertujuan untuk membuat pemodelan berstatistik dan meramalkan nilai pemboleh ubah bersandar berdasarkan nilai pemboleh ubah tidak bersandar, serta menilai kekuatan hubungannya (Montgomery *et al.*, 2012). Penyelidikan yang berkaitan dengan bidang perubatan, analisis regresi biasanya digunakan untuk menentukan diagnosis yang tepat (Palta, 2021). Menurut Ross (2020), terdapat dua jenis model

regresi yang sering diguna pakai, iaitu model regresi linear (yang mana hubungan di antara dua pemboleh ubah boleh diwakili oleh suatu persamaan garis lurus) dan model regresi yang tidak linear (yang mana hubungan antara dua pemboleh ubah tidak boleh diwakili oleh suatu persamaan garis lurus). Umumnya, model regresi linear adalah yang paling banyak digunakan dalam pelbagai kajian penyelidikan dan ianya merentas pelbagai disiplin ilmu. Hal ini kerana, hubungan seperti ini mudah diinterpretasikan serta mudah untuk dibentuk menerusi pemboleh ubah yang dikaji (Ross, 2020). Regresi linear berganda boleh dilihat sebagai suatu pendekatan lanjutan berbanding dengan kaedah regresi linear mudah. Hal ini kerana ia menggabungkan lebih daripada satu pemboleh ubah peramal (Ross, 2020). Teknik ini biasanya digunakan dalam penyelidikan yang melibatkan data yang bertaburan secara linear di antara satu pemboleh ubah sambutan dengan beberapa pemboleh ubah peramal. Sebagai contoh, dalam kajian yang dijalankan oleh Nawi *et al.* (2018), yang mana beliau menggunakan model regresi linear berganda untuk meramalkan bacaan aras gula (pemboleh ubah bersandar) dalam kalangan pesakit kencing manis dengan pemboleh ubah tidak bersandar termasuk indeks jisim badan (*BMI*), berat badan, paras kolesterol, ketinggian, dan tekanan darah sistolik.

Ahmad *et al.* (2016) telah mencadangkan bahawa penggunaan metodologi hibrid adalah sangat berkesan untuk pemodelan secara berstatistik, walaupun kaedah tersebut masih belum diterokai sepenuhnya. Pendekatan statistik secara hibrid mempunyai potensi yang sangat baik untuk menghasilkan suatu keputusan yang sangat tepat dengan menggabungkan pelbagai kaedah. Jika dibandingkan dengan pendekatan analisis secara tunggal, metodologi hibrid secara konsisten menghasilkan keputusan yang lebih tepat dan mengagumkan (Ahmad *et al.*, 2016). Kajian oleh Adnan *et al.* (2023) yang melibatkan 200 data forensik simulasi, metodologi hibrid yang

menggabungkan model regresi linear (*ALRM*) yang digunakan dan perceptron berbilang lapisan (*MLP*) menunjukkan suatu keputusan yang amat memberangsangkan. Oleh hal yang demikian, satu inisiatif telah diambil untuk meneruskan lagi kajian tersebut dengan menggabungkan model regresi linear berganda dengan model rangkaian neural buatan serta mempertimbangkannya dengan membangunkan “Antara Muka Pengguna Grafik - *Graphical User Interface (GUI)*” dan diaplikasikannya dalam sains kesihatan yang dinamakan sebagai model *LiReNN*. Selain itu, teknik ‘butstrap’ digunakan untuk meningkatkan kebolehamalan dan ketepatan model regresi. Penyelidikan ini amatlah bernilai kepada penyelidik, khususnya dari segi pengesahan model yang diperoleh serta kebolehamannya membuat peramalan (Ahmad *et al.*, 2020). Tambahan pula, kaedah antara muka pengguna grafik (*GUI*) turut dimasukkan ke dalam kajian ini, kaedah ini adalah mesra pengguna, ianya dibangunkan untuk meningkatkan kebolehcapaian dan kebolegunaan analisis. Justeru, aplikasi *GUI* yang digunakan dalam kajian ini berperanan sebagai elemen yang membenarkan pengguna berkomunikasi dan berinteraksi dengan sistem pengoperasian secara terus dalam menghasilkan keputusan yang baik (Rahmalia, 2021).

Pembangunan metodologi yang melibatkan gabungan model linear dan model rangkaian neural buatan dalam sains kesihatan merupakan pendekatan yang menarik dan mempunyai potensi yang sangat tinggi dalam bidang penyelidikan (Ghazali *et al.*, 2021; Bharti *et al.*, 2023). Pendekatan ini akan memanfaatkan kelebihan pada kedua-dua jenis model untuk meningkatkan pemodelan dan ramalan dalam bidang sains kesihatan. Berikut adalah beberapa langkah yang boleh diambil untuk membangunkan metodologi seperti berikut:

- (a) Proses ini bermula dengan mengumpul dan menyusun data kesihatan yang berkaitan.

- (b) Data ini kemudiannya dianalisis menggunakan kaedah statistik untuk memahami hubungan antara pemboleh ubah.
- (c) Model linear digunakan untuk memodelkan hubungan di antara pemboleh ubah bersandar dan tidak bersandar, memberikan pemahaman awal tentang faktor-faktor yang mempengaruhi masalah kesihatan.
- (d) Rangkaian neural buatan dilatih menggunakan data yang sama untuk menentukan hubungan yang lebih kompleks.
- (e) *Output* daripada kedua-dua model linear dan rangkaian neural digabungkan menggunakan pelbagai kaedah.
- (f) Prestasi model gabungan dinilai dengan menggunakan metrik seperti ketepatan dan teknik pengesahan silang.
- (g) Sekiranya terdapat keperluan, model ini haruslah diperhalusi dan dioptimumkan.
- (h) Akhirnya, setelah metodologi telah diuji dan dioptimumkan. Maka metodologi ini boleh berkeupayaan untuk melaksanakan analisis dalam situasi yang sebenar.

Selain daripada itu, metodologi ini juga turut menjaga isu-isu privasi dan keselamatan data yang berkaitan dengan penggunaan maklumat kesihatan. Dengan cara ini, metodologi yang menggabungkan model linear dan rangkaian neural buatan dapat memberikan sumbangan yang bernilai dalam pembangunan sains kesihatan dan memberikan pemahaman yang lebih baik mengenai masalah kesihatan serta kemungkinan penyelesaiannya.

#### **1.4.1 Keperluan pemodelan secara model regresi linear berintegrasi dengan model rangkaian neural dalam bidang sains kesihatan**

Metodologi yang menggabungkan model linear dan rangkaian neural buatan dapat memberikan manfaat yang signifikan dalam pemodelan dan ramalan dalam sains

kesihatan. Pendekatan ini dapat membantu dalam menganalisis dan meramalkan pelbagai masalah kesihatan, seperti diagnosis penyakit, penilaian tahap risiko kesihatan, atau peramalan hasil rawatan. Selain itu, pemodelan regresi yang diintegrasikan dengan rangkaian neural memainkan peranan penting dalam bidang sains kesihatan. Pemodelan regresi adalah merupakan teknik statistik yang digunakan untuk mempelajari hubungan antara satu atau lebih pemboleh ubah tidak bersandar (*independent variable*) dengan pemboleh ubah bersandar (*dependent variable*) dalam bentuk fungsi statistik.

Tambahan pula, dalam bidang sains kesihatan, pemodelan regresi berintegrasi dengan rangkaian neural memungkinkan analisis yang lebih kompleks dan pemahaman yang lebih mendalam tentang hubungan antara pemboleh ubah klinikal, genetik, lingkungan, dan pemboleh ubah kesihatan lainnya. Dengan menggunakan rangkaian neural sebagai alat pemodelan, kelebihan-kelebihan berikut dapat dicapai seperti rangkaian neural mempunyai keupayaan untuk mengenali corak kompleks dalam data kesihatan, membolehkan pemahaman yang lebih mendalam tentang faktor yang mempengaruhi kesihatan, juga boleh membuat ramalan yang lebih tepat tentang kejadian penyakit, tindak balas rawatan dan hasil jangka panjang dengan memodelkan pemboleh ubah yang berkaitan. Selain itu, rangkaian neural boleh membantu mengoptimalkan pengecaman terhadap faktor kajian yang dikaji, pengurusan rawatan dan kesihatan dengan memahami cara faktor yang berbeza berinteraksi. Akhir sekali, pembentukan kepada pengetahuan baharu dalam bidang sains kesihatan dengan menganalisis set data yang besar dan kompleks, yang membawa kepada penemuan baharu dan strategi pencegahan dan rawatan yang lebih baik.

Kesimpulannya, integrasi pemodelan regresi linear dengan rangkaian neural memberikan pendekatan yang kuat dan inovatif untuk menganalisis data kesihatan dan

memahami punca yang mempengaruhi kesihatan dan penyakit. Ini membawa potensi untuk meningkatkan diagnosis, pengubatan, pengurusan, dan pengetahuan di bidang sains kesihatan secara keseluruhan.

### **1.5 Pernyataan masalah**

Kajian terdahulu memberi tumpuan kepada satu teknik tunggal sahaja, yang mempunyai potensi terhad untuk menyelesaikan masalah yang berkaitan dengan sains kesihatan. Menurut hasil daripada kajian yang lepas, kaedah tunggal yang dicadangkan kurang komprehensif. Justeru, keperluan untuk mendapatkan suatu model yang lebih tepat amatlah diperlukan. Oleh itu, model statistik yang menggabungkan teknik seperti model *LiReNN* akan membolehkan anggaran yang lebih tepat untuk tujuan pemodelan. Model regresi linear berganda (*MLR*) dengan rangkaian neural hadapan suapan berbilang lapisan (*MLFFNN*) menjadi fokus utama kajian ini kerana dipercayai bahawa gabungan model ini boleh digunakan dalam pelbagai keadaan bagi tujuan permodelan. Malahan, model gabungan yang dicadangkan dapat digunakan untuk mengkaji kewujudan “sebab-akibat” dan mengaitkannya dengan beberapa faktor berpotensi yang menyumbang kepada masalah yang dikaji.

Terdapat suatu jurang fungsi di antara metodologi yang dikemukakan. Disebabkan hal demikian, membangunkan serta mengaplikasikan kaedah model *LiReNN* akan melibatkan suatu perubahan pengubahsuaian yang signifikan kepada kedua-dua kaedah untuk mengatasi jurang penyelidikan tersebut. Kajian ini juga bertujuan untuk meneroka bagaimana penggunaan gabungan model *ANN* dengan regresi linear dapat memberikan pemahaman yang lebih mendalam tentang hubungan antara pemboleh ubah dalam sains kesihatan. Penggunaan teknik visualisasi data kompleks dapat membantu dalam memahami interaksi antara pemboleh ubah yang

ditemui melalui model gabungan ini, memberikan wawasan baru tentang faktor-faktor yang mempengaruhi keadaan kesihatan. Keberkesanan model *LiReNN* dinilai dengan menggunakan bacaan Ralat Kuasa Dua Min (*Mean Square Error-MSE*) dan peratusan ketepatan.

Gabungan teknik secara hibrid menerusi kaedah statistik masih belum diterokai sepenuhnya, terutamanya dalam kajian mengenai penyakit kronik di Malaysia. Justeru, kaedah yang dicadangkan ini mempunyai potensi yang tinggi untuk terus berkembang, malah ianya juga mempunyai keupayaan jauh lebih baik untuk aplikasikan dalam pelbagai bidang terutamanya untuk menyelesaikan masalah berkaitan dengan sains kesihatan. Tambahan pula, kajian perlu dilakukan untuk menguji dan memvalidasi keberkesanan gabungan model *ANN* dengan regresi linear dalam menghasilkan ramalan yang sah dan berguna dalam sains kesihatan. Penelitian juga boleh menumpukan kepada pembangunan kaedah penilaian yang lebih canggih untuk menilai kesesuaian model gabungan ini dalam pelbagai situasi dalam bidang sains kesihatan.

Kekurangan aplikasi yang menggunakan perisian "Antara Muka Pengguna Grafik (*GUI*)" telah menjadi fokus perhatian dalam pembangunan model *LiReNN*. Ini disebabkan oleh kepentingan untuk membuat perisian lebih mudah diakses oleh pengguna yang mungkin tidak memiliki kepakaran teknikal yang tinggi. Oleh itu, dalam konteks ini, *GUI* telah direkabentuk khusus untuk memastikan bahawa pengguna dapat berinteraksi dengan perisian secara efisien tanpa memerlukan pengetahuan teknikal yang mendalam.

Selain itu, *GUI* telah diintegrasikan ke dalam program yang merangkumi kaedah model regresi linear dengan rangkaian neural buatan. Penekanan diberikan kepada



unsur kecekapan, tahap kompetensi model, membuat peramalan, dan mengesahkan ketepatan model *LiReNN*. Ini bermakna bahawa *GUI* tidak hanya menyediakan antarmuka yang intuitif untuk berinteraksi dengan perisian, tetapi juga memfasilitasi proses analisis dan pemodelan dengan menggunakan teknik regresi linear dan rangkaian neural. Ini membolehkan pengguna untuk dengan mudah menavigasi melalui model *LiReNN*, menganalisis hasil, dan membuat keputusan berdasarkan hasil ramalan dan penilaian ketepatan model.

## **1.6 Rasional penyelidikan**

Dalam bidang sains perubatan dan kesihatan, keperluan ketara dalam metodologi penyelidikan yang teguh menjadi fokus utama, terutamanya yang berkaitan dengan penggabungan kaedah atau prosedur. Sementara itu, kerangka konsep bagi kajian ini yang menggunakan model *LiReNN* ditunjukkan dalam Rajah 1.1. Tambahan pula, permintaan terhadap model yang sangat tepat dan hasil yang dioptimumkan semakin meningkat pada masa kini. Oleh itu, kajian ini memberi tumpuan kepada penggabungan model yang sangat tepat dengan penekanan yang kuat pada kebolehpercayaan, dengan matlamat untuk memaksimumkan kebolegunaan hasil kajian. Salah satu kelemahan besar dalam perkembangan metodologi berdasarkan penyelidikan yang lepas adalah ketiadaan mekanisme pengesahan model serta ketiadaan *GUI* yang komprehensif yang membantu penyelidik untuk memudahkan capaian analisis. Di samping itu, adalah penting untuk mencipta penyelesaian bagi mengatasi kekurangan ini. Cabaran ini boleh diatasi melalui penggabungan metodologi sedia ada, serta dengan penyempurnaan atau pengubahsuaian kaedah semasa.

## **1.7 Hipotesis penyelidikan**

Penyelidikan ini penting untuk pembangunan asas metodologi pengkomputeran statistik. Hipotesis penyelidikan adalah seperti berikut:

- i. Gabungan kaedah model regresi linear dengan rangkaian neural dapat meningkatkan keberkesanan dan kecekapan analisis.
- ii. Kaedah gabungan ini dapat digunakan untuk membuat peramalan serta menentusahkan kebaikan penyesuaian model yang dibina secara berkesan.
- iii. Antara Muka Pengguna Grafik (*GUI*) meningkatkan kebolegunaan serta kemudahan analisis.

## **1.8 Objektif penyelidikan umum**

Objektif penyelidikan menerangkan hala tuju dan matlamat penyelidikan. Kajian ini memberi tumpuan utama kepada proses pembangunan kaedah rangkaian neural dengan menggabungkan model regresi linear secara komprehensif berteraskan kepada statistik komputasi dengan menggunakan perisian *RStudio* (bahasa pengaturcaraan *R*) dan *Eclipse* (bahasa pengaturcaraan *Java*) untuk membangunkan antara muka pengguna grafik (*Graphical User Interface-GUI*).

### **1.8.1 Objektif penyelidikan khusus**

Objektif khusus kajian ini merangkumi objektif-objektif yang berikut:

- i. Membangunkan kaedah model regresi linear dengan rangkaian neural yang menitikberatkan unsur kecekapan dan tahap kompetensi model.
- ii. Membuat peramalan serta menentusahkan ketepatan model yang dibina.
- iii. Membangunkan rekabentuk Antara Muka Pengguna Grafik (*GUI*) untuk objektif (i) dan (ii) bagi meningkatkan kebolegunaan serta kemudahan analisis.

## 1.9 Persoalan penyelidikan

Persoalan kajian juga dinyatakan bagi mencapai matlamat yang sejajar dengan tajuk kajian ini. Persoalan kajian yang mendasari kajian penyelidikan dapat diringkaskan seperti berikut:

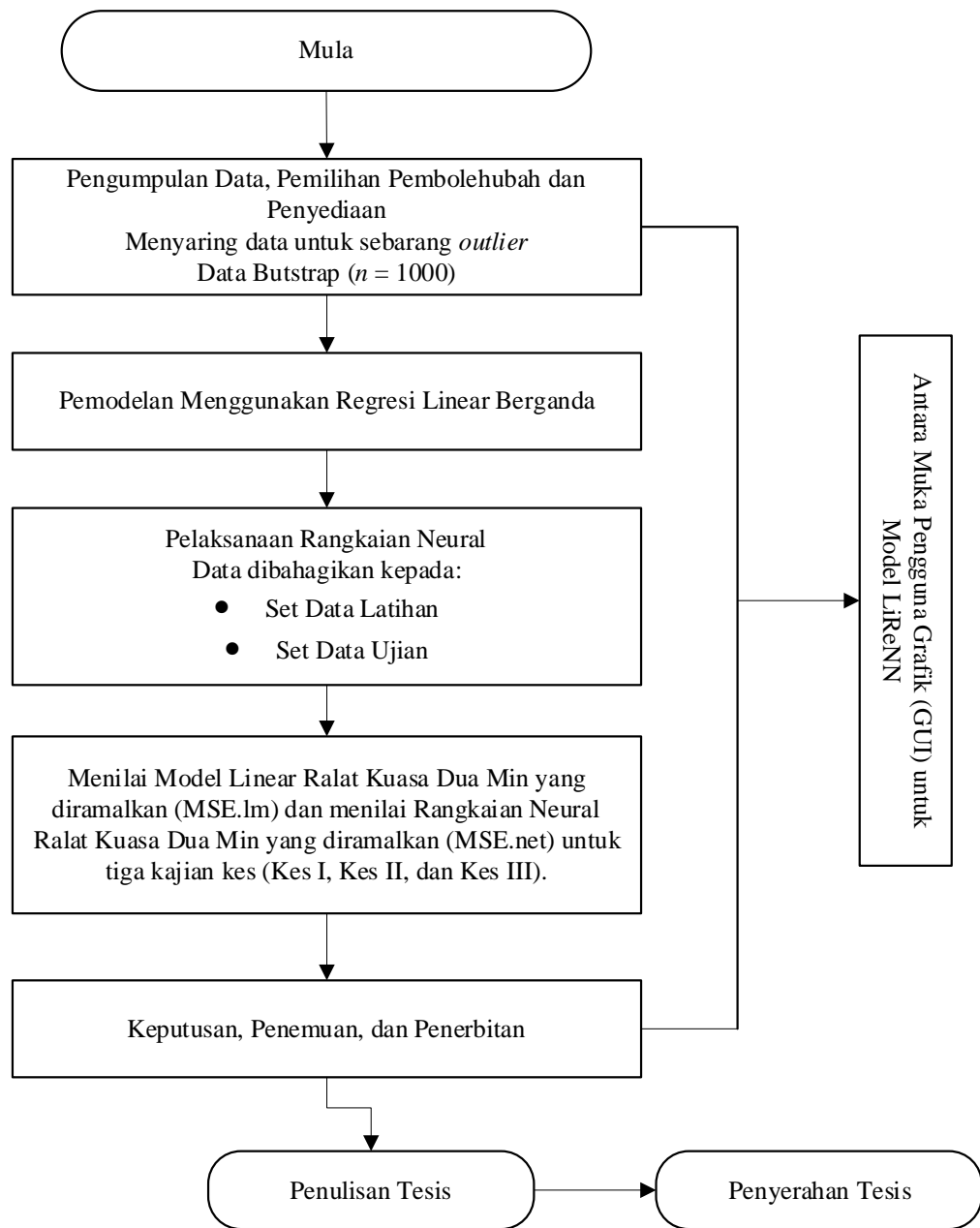
- (a) Adakah kaedah yang dibangunkan secara regresi linear dengan rangkaian neural adalah cekap dan mempunyai tahap kompetensi yang tinggi?
- (b) Bagaimanakah peramalan dibuat dan menentukan ketepatan model yang dibina?
- (c) Adakah *GUI* yang direkabentuk berdasarkan objektif (i) dan (ii) berhasil?

## 1.10 Skop penyelidikan kajian

Penyelidikan ini dihadkan kepada berikut:

- (a) Skop utama kajian ini adalah melibatkan pembangunan metodologi secara model regresi linear dengan model rangkaian neural yang menitikberatkan unsur kecekapan dan tahap kompetensi model.
- (b) Pengujian terhadap pembangunan metodologi dilaksanakan menerusi pengujian kesesuaian model, dan kestabilan reka bentuk rangkaian neural.
- (c) Selain daripada itu, kajian ini menggunakan gabungan perisian *RStudio* dalam pembentukan bahasa pengaturcaraan *R*. Manakala, *Java* akan digunakan untuk membangunkan *GUI* dengan menyatukan bahasa pengaturcaraan tersebut dalam sistem yang dibina menggunakan perisian *Eclipse*.
- (d) Skop pembangunan metodologi dalam kajian ini hanya berteraskan kepada bidang sains kesihatan yang merangkumi ke semua kriteria yang telah ditetapkan. Disamping itu, skop kajian mengenai penugasan data hanya tertumpu pada set data daripada pesakit diabetes mellitus jenis 2.

## 1.11 Kerangka penyelidikan keseluruhan



Rajah 1.1: Kerangka kajian secara keseluruhan

Kerangka kajian secara keseluruhan ini bertujuan untuk mewujudkan pendekatan yang komprehensif dengan menggabungkan kaedah regresi linear berganda dengan rangkaian neural untuk kes sains kesihatan (kajian kes I, kajian kes II, dan kajian kes III), seperti yang ditunjukkan dalam Rajah 1.1.

## 1.12 Sumbangan kajian

Kajian ini dapat memberikan manfaat dalam sains kesihatan dan pembangunan metodologi menerusi kaedah model linear berganda dengan rangkaian neural buatan yang menitikberatkan unsur kecekapan dan tahap kompetensi model. Trend teknologi yang semakin berkembang dari semasa ke semasa memerlukan kajian berterusan.

Sumbangan kajian ini boleh diringkaskan kepada tiga bahagian utama. Sumbangan pertama berfokus pada pembangunan metodologi yang terintegrasi yang boleh digunakan dalam kes-kes perubatan dan kesihatan untuk meramalkan masa depan. Kajian ini secara signifikan menyumbang kepada pembangunan metodologi yang terdiri daripada dua gabungan kaedah yang dicadangkan iaitu regresi linear berganda dan rangkaian neural buatan. Dengan menggabungkan kedua-dua kaedah ini, kajian ini memberi sumbangan yang penting dalam meningkatkan keberkesanan dan ketepatan dalam meramalkan kejadian-kejadian penting dalam bidang perubatan dan kesihatan.

Sumbangan kedua adalah dalam konteks sintaks *R* untuk alat statistik. Penggunaan sintaks *R* membolehkan penyelidik untuk menyesuaikan kod secara spesifik untuk keperluan analisis data. Dengan menggunakan sintaks *R*, penyelidik dapat menyesuaikan dan mengoptimumkan proses analisis statistik untuk meningkatkan prestasi. Disamping itu, penyelidik dapat menyesuaikan kod untuk memilih dan mengatur pembolehubah yang paling relevan atau penting dalam kajian mereka (Wickham, 2015). Tambahan pula, sintaks *R* sering menghasilkan hasil yang terperinci dan mudah difahami. Ini bermakna penyelidik dapat melihat dengan jelas *output* analisis statistik dan membuat keputusan yang tepat berdasarkan hasil yang diperolehi. Hasil yang terperinci juga memudahkan penyelidik dalam membuat kesimpulan yang

berasaskan bukti dan merumuskan interpretasi yang tepat dari data yang dianalisis (Venables *et al.*, 2023). Dengan demikian, sumbangan sintaks *R* dalam kajian ini tidak hanya meningkatkan keseluruhan prestasi analisis data, tetapi juga memperkuat proses pengambilan keputusan penyelidik.

Sumbangan ketiga, kajian ini menyumbang kepada literatur sains yang sedia ada dan membolehkan penyelidik lain meneruskan usaha penyelidikan masa depan. Ini akan membantu penyelidik baru memahami cabaran yang berkaitan dengan mereka bentuk metodologi baru atau menjalankan penyelidikan. Selain itu, kajian ini boleh mencadangkan kaedah yang lebih baik untuk tujuan pemodelan dengan ramalan dan ketepatan yang tinggi.

### **1.13 Keterbatasan kajian penyelidikan**

Keterbatasan kajian penyelidikan merujuk kepada had atau kelemahan yang terdapat dalam penyelidikan saintifik. Beberapa keterbatasan kajian penyelidikan adalah seperti berikut:

a) Butstrap

- Kaedah butstrap digunakan untuk menilai ketepatan dan kesesuaian model statistik, serta untuk menghasilkan ramalan dan menganggar set data. Oleh itu, kajian ini hanya melibatkan pensampelan semula kes data yang bertujuan untuk memperhalusi parameter yang dianggarkan.

b) Regresi Linear Berganda (*MLR*)

- Hanya menggunakan kaedah masuk, semua pemboleh ubah yang disertakan sudah bersih dan mempunyai kepentingan yang signifikan secara klinikal daripada literatur sebelumnya.

c) Rangkaian Neural Buatan/Rangkaian Neural Hadapan Suapan Berbilang lapisan (*MLFFNN*)

- Metodologi ini hanya digunakan untuk mengesahkan faktor yang diperolehi dalam (b), mengira ketepatan, dan menilai kebolehamalan metodologi. Keputusan yang diramalkan (data ujian) akan dibandingkan dengan nilai sebenar yang menunjukkan keberkesanan pendekatan yang dicadangkan.

d) Antara Muka Pengguna Grafik (*GUI*)

- Hanya membangunkan *GUI* bagi aplikasi model *LiReNN*.

Selain itu, keterbatasan kajian juga berkaitan dengan pemilihan perisian yang digunakan. Memandangkan *R* ialah platform perisian yang percuma dan mudah digunakan, semua metodologi yang dirangka akan dilaksanakan sepenuhnya dalam aplikasi ini. Kajian ini menggunakan data sekunder hasil daripada kajian penyelidikan yang lepas. Kajian kes I yang dikaji adalah mengenai pemodelan bacaan aras kreatinin dalam kalangan pesakit dislipidemia dengan diabetes mellitus jenis 2, manakala kajian kes II adalah pemodelan terhadap indeks jisim badan dalam kalangan pesakit diabetes mellitus jenis 2, dan kajian kes III adalah pemodelan aras lipoprotein berketumpatan rendah dalam kalangan pesakit diabetes mellitus jenis 2. Ke-semua sumber data yang digunakan adalah diperolehi daripada klinik pesakit luar di Hospital Universiti Sains Malaysia (USM), Kelantan, Malaysia.

#### **1.14 Organisasi tesis**

Tesis ini terdiri daripada enam bab yang disusun dengan teliti, setiap satu disusun dengan sempurna untuk menunjukkan naratif yang menarik. Kandungan Bab 1 merangkumi pengenalan terhadap kajian yang dijalankan serta membincangkan aspek

lain seperti latar belakang dan perkembangan kajian, pernyataan masalah, rasional kajian, hipotesis kajian, objektif kajian, persoalan kajian, skop kajian, kerangka kajian secara keseluruhan, sumbangan kajian, dan keterbatasan kajian. Bab 2 pula memberikan penjelasan yang lebih terperinci tentang sorotan kesusasteraan yang berkaitan dengan kaedah-kaedah seperti regresi linear, butstrap, rangkaian neural, dan *GUI*.

Bab 3 menjelaskan tentang perancangan kajian, lokasi kajian, dan tempoh penyelidikan. Penerangan tentang kaedah pensampelan tidak digunakan dalam kajian ini. Bab ini juga merangkumi proses pengumpulan data dan pemilihan pemboleh ubah untuk setiap kajian kes. Selain itu, langkah-langkah dalam membina algoritma untuk setiap model menggunakan bahasa pengaturcaraan *R* turut disertakan untuk memudahkan pemahaman penggunaan *R* dalam melaksanakan analisis model regresi linear dengan rangkaian neural buatan. Tambahan pula, dalam bab ini juga menunjukkan langkah-langkah mereka bentuk antara muka pengguna grafik (*GUI*) dengan menggunakan perisian *Eclipse* serta bahasa pengaturcaraan *Java*.

Bab 4 mengemukakan hasil kajian melalui penggunaan model regresi linear yang dibina dengan menggunakan beberapa kaedah utama seperti kaedah regresi linear, butstrap, dan rangkaian neural buatan. Setiap hasil kajian yang diperoleh akan dijelaskan secara terperinci untuk memudahkan pemahaman tentang penyelidikan yang dilakukan.

Bab 5 merupakan bab yang memfokuskan kepada perbincangan dan ringkasan mengenai penyelidikan yang dilakukan. Aspek yang diberikan penekanan dan tumpuan utama ialah kebaruan model *LiReNN* yang dibina dengan menggunakan pendekatan alternatif. Bab 6 adalah bab terakhir dalam kajian yang memberikan penjelasan



mengenai beberapa perkara penting seperti batasan kajian, manfaat hasil penyelidikan, dan cadangan untuk kajian pada masa hadapan. Ia menggariskan had-had penyelidikan, menyatakan faedah hasil kajian, dan mencadangkan arah untuk penyelidikan lanjutan.

## **BAB 2**

### **SOROTAN KESUSASTERAAN**

#### **2.1 Pengenalan bab**

Bab 2 menyajikan tentang kesusasteraan lepas yang berkaitan dengan kajian yang dijalankan ini. Hal ini bertujuan untuk mendapatkan suatu gambaran yang menyeluruh mengenai kaedah yang dikaji di samping untuk mengumpul idea-idea dan merumuskan penemuan utama mengenai penggunaan kaedah statistik dalam bidang sains kesihatan. Seterusnya pada bahagian 2.2, ianya membincangkan tentang sejarah dan perkembangan rangkaian neural yang merangkumi pendekatan rangkaian neural dalam bidang sains kesihatan, tinjauan berkaitan dengan aplikasi, dan serta kelebihan menggunakan model rangkaian neural. Sementara itu, bahagian 2.3 pula menerangkan tentang sejarah dan perkembangan model regresi linear serta tinjauan terhadap aplikasi regresi linear dan kelebihannya. Seterusnya, bahagian 2.4 menjelaskan tentang kepentingan kaedah persampelan berulang, manakala bahagian 2.5 membahas antara muka pengguna grafik. Disamping itu, bahagian 2.6 menjelaskan mengenai diabetes mellitus, dan bahagian 2.7 menerangkan secara ringkas ketiga-tiga kajian kes menggunakan kaedah statistik yang digunakan sebelum ini. Bab ini juga merangkumi batasan kajian lepas serta rumusan akhir.

#### **2.2 Bahagian I: Sejarah dan perkembangan model rangkaian neural**

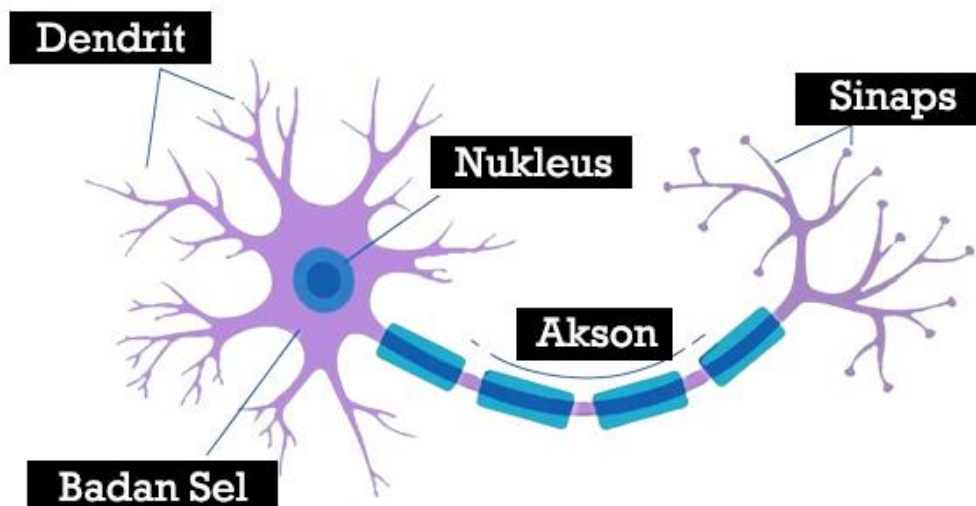
Model rangkaian neural adalah merupakan satu teknologi yang telah wujud sejak lama, namun ianya menjadi semakin popular dan berkembang dengan pesat dalam beberapa tahun kebelakangan ini. Teknologi ini terinspirasi oleh struktur otak manusia dan diciptakan untuk memproses dan menganalisis data secara efektif, termasuklah set data yang kompleks dan besar (won *et al.*, 2016). Model rangkaian neural sebenarnya berkonsepkan corak pemikiran manusia yang berasaskan gabungan set saraf yang

berkait rapat. Otak manusia pada hakikatnya terdiri daripada satu set saraf yang berkait rapat, atau unit pemprosesan maklumat asas, yang dipanggil neuron. Menurut Shepherd (2004) dan Worden *et al.* (2023), otak manusia merangkumi kira-kira 10 bilion neuron dan 60 trilion sinaps yang menakjubkan, yang memudahkan komunikasi antara neuron ini. Dengan bergabungnya beberapa neuron secara serentak, otak dapat menjalankan fungsi-fungsi dengan jauh lebih pantas berbanding sistem komputer terkini (Michael, 2005; Worden *et al.*, 2023).

Tambahan lagi, Aini *et al.* (1993) dan Heddam (2014) menjelaskan bahawa gabungan berjuta-juta rangkaian neuron atau saraf adalah terdiri daripada sistem saraf otak manusia dan haiwan, ia juga dipanggil sebagai rangkaian neural atau saraf biologi. Di samping itu, rangkaian neural buatan/saraf tiruan atau rangkaian saraf ialah rangkaian yang dicipta berdasarkan model otak manusia yang terdiri daripada rangkaian sel neuron yang boleh dimodelkan sebagai peranti dua dimensi (model McCulloch-Pitts) dengan cara yang paling mudah (Walczak, 2019; Chen *et al.*, 2023). Setiap rangkaian neuron atau saraf terdiri daripada sel badan (yang mengandungi nukleus sel), akson (bahagian awal laluan sambungan), sinapsis (pemancar isyarat impuls saraf), dan dendrit (laluan penghubung) (Siddharth, 2021).

Pada tahun 2023, seorang penyelidik bernama Gill telah membincangkan secara terperinci bahagian neuron dan fungsinya. Komponen asas sel saraf dalam otak manusia terdiri daripada empat bahagian iaitu dendrit, soma (badan sel), akson, dan sinaps, seperti yang ditunjukkan dalam Rajah 2.1. Peranan utama dendrit adalah untuk menerima dan mengintegrasikan isyarat yang dihantar oleh neuron lain. Sel soma, juga dikenali sebagai badan sel yang melaksanakan fungsi menggabungkan dan menyepadukan semua isyarat masuk yang diterima oleh neuron untuk menjana *input*

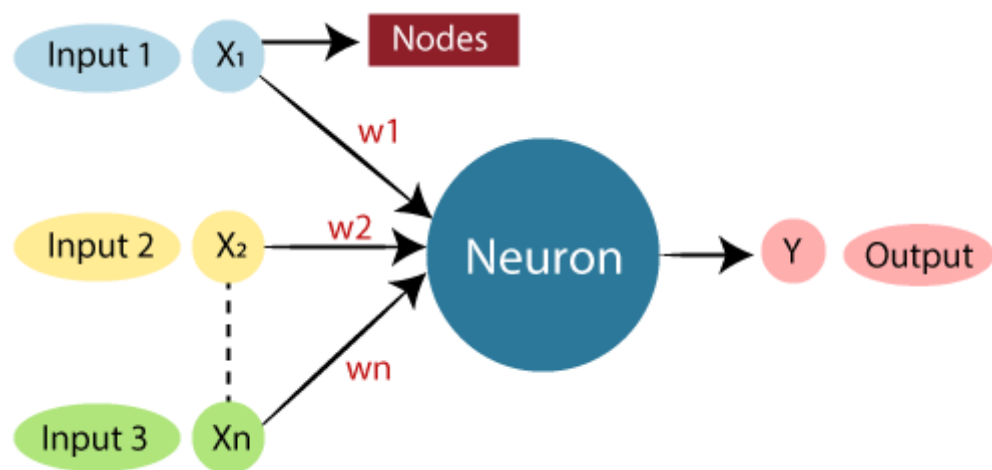
keseluruhan. Selain itu, struktur akson bertanggungjawab untuk penghantaran *input* ini ke neuron lain apabila ia melepasi nilai ambang tertentu. Sinaps yang berfungsi sebagai persimpangan yang mana satu neuron berhubung dengan yang lain, memainkan peranan penting dalam menghantar sesuatu isyarat. Penghantaran isyarat adalah melalui sambungan ini ditentukan oleh pemberat sinaptik, semakin berat sesuatu sinaptik semakin kuat sambungan tersebut. Ikatan tersebut boleh menjadi semakin lemah atau boleh menjadi semakin kuat. Oleh itu, rangkaian neural lazimnya terdiri daripada rangkaian besar yang mempunyai berbilion neuron yang dihubungkan dengan bertrilion sambungan (Gill, 2023).



Rajah 2.1: Rangkaian neuron biologi (Sumber: Gill, 2023)

Rangkaian neural biologi (*BNN*) menjadi sumber inspirasi utama untuk membina rangkaian neural buatan bagi meniru mereka agar mesin dapat menjalankan tugas yang kompleks dan berfikir seperti manusia. Neuron merupakan unit utama dalam rangkaian neural (Siddharth, 2021). Selain itu, istilah “Rangkaian Neural Buatan (RNB)” berasal daripada rangkaian neural biologi yang membentuk struktur otak manusia. Hal ini menerangkan bahawa otak manusia yang mempunyai neuron yang

bersambung antara satu sama lain, RNB juga mempunyai neuron yang bersambung antara satu sama lain dalam pelbagai lapisan rangkaian. Neuron-neuron ini dikenali sebagai nod. Dalam RNB, *input* diwakili oleh dendrit, nod diwakili oleh nukleus sel, pemberat diwakili oleh sinaps, dan *output* diwakili oleh akson (Kim *et al.*, 2013; Sharifi *et al.*, 2019).



Rajah 2.2: Struktur rangkaian neural buatan (Sumber: Siddharth, 2021)

Sejarah rangkaian neural bermula pada tahun 1943 ketika dua ahli matematik, Warren McCulloch dan Walter Pitts, memperkenalkan model matematik untuk neuron yang merupakan satu unit dasar yang membentuk rangkaian neural. Pada awal tahun 1950, ilmuwan yang terkenal iaitu Frank Rosenblatt telah memperkenalkan perceptron, iaitu sejenis rangkaian neural yang dapat digunakan untuk pengenalan pola. Namun, perkembangan teknologi komputer pada tahun 1970-an sehingga 1980-an membawa kepada peningkatan minat dalam perkembangan rangkaian neural. Seterusnya, pada tahun 2000, kemajuan teknologi sekali lagi meningkat dan ke semua peranti keras dan perisian komputer, rangkaian neural kembali menjadi fokus dalam bidang sains komputer dan pembelajaran mesin (Walczak, 2019). Tidak ketinggalan juga model rangkaian neural turut mengalami peningkatan. Pada detik ini, terdapat banyak jenis