

فصل چهارم

مواد داده

فصل ۴

مواد داده

در این فصل داده های مورد استفاده در این پایان نامه ارائه گردیده است.

۱.۴ مجموعه

مجموعه داده مورد استفاده در این پایان نامه توسط پروژه تحقیقاتی تولدهای ایمن جمع آوری شده است، که یک همکاری تحقیقاتی با شرکای شامل، اما نه محدود به، دانشگاه استوانگر، بهداشت جهانی Laerdal، و بیمارستان های شریک در تانزانیا است. داده ها بین اکتبر ۲۰۱۵ و ژوئن ۲۰۱۸ جمع آوری شد و قبل از انتقال به تحقیق ناشناس شدند.

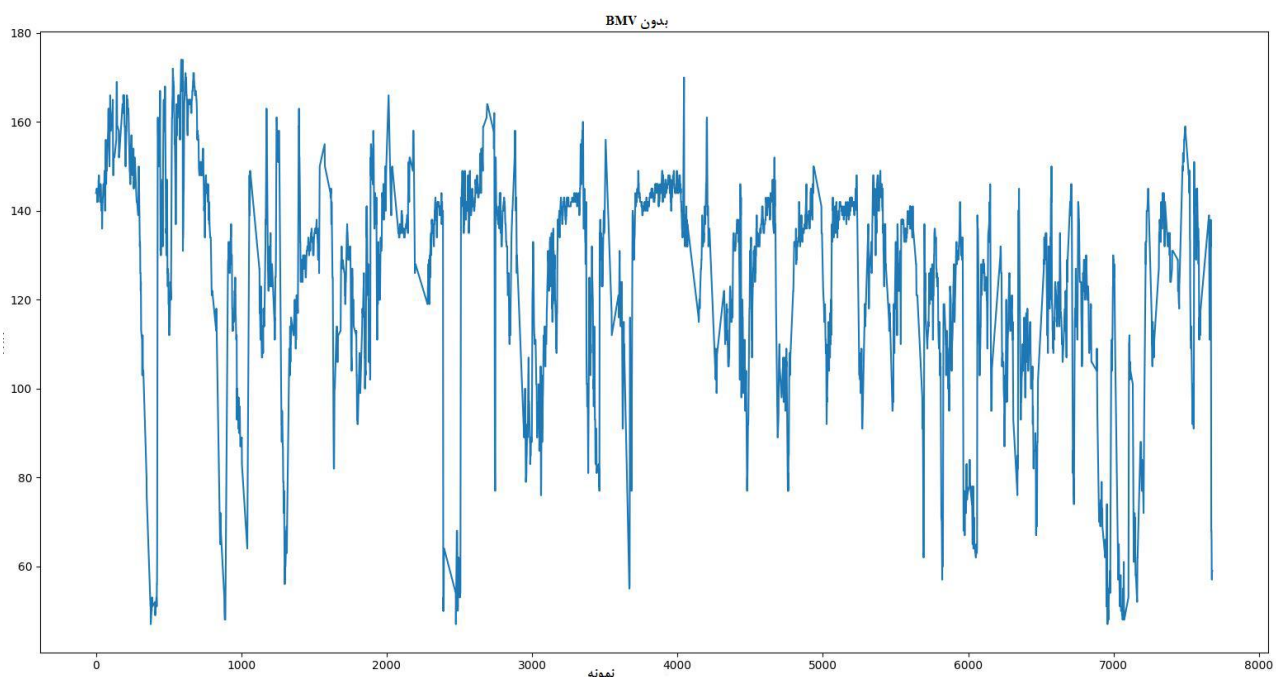
۲.۴ مجموعه داده

مجموعه داده مورد استفاده در این مطالعه از ۳۱۱۱ تولد جمع آوری شده است. علاوه بر سیگنال های FHR، حاوی اطلاعات بالینی اضافی بود که شرح زایمان توسط دستیاران تحقیقاتی تعیین شده حاضر در بخش زایمان را شرح می دهد. این اطلاعات شامل، اما نه محدود به، نتیجه پس از ۳۰ دقیقه و نتیجه پس از ۲۴ ساعت است.

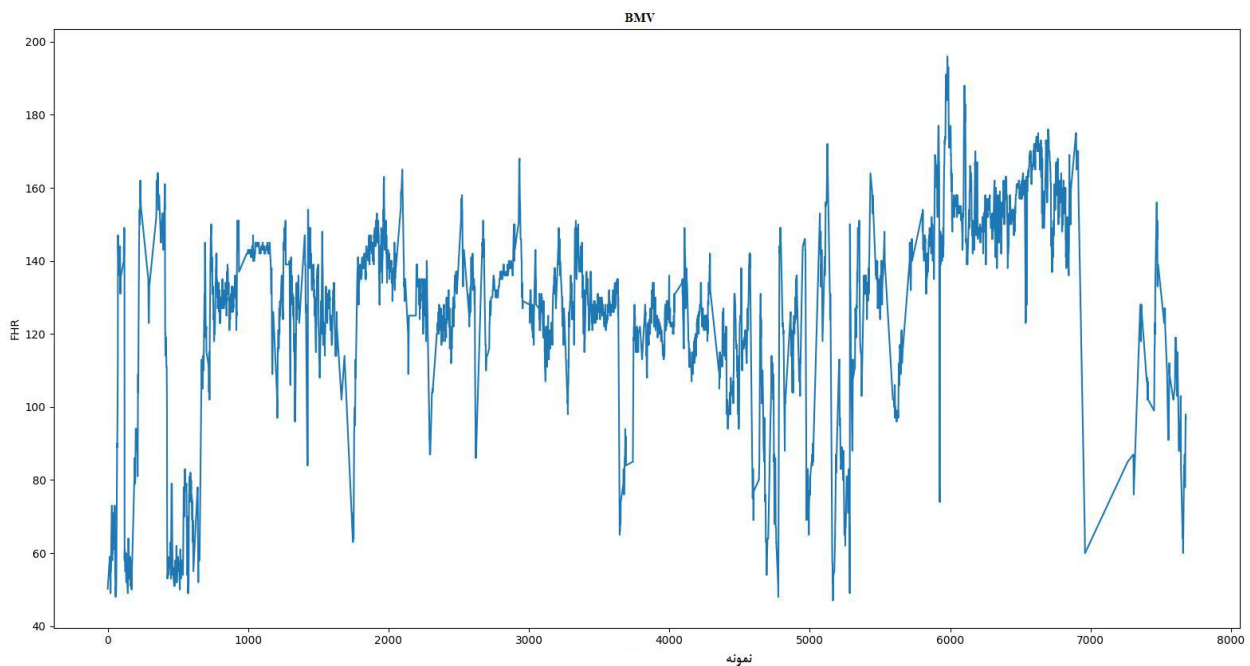
FHR اندازه گیری شده مستعد ایجاد اختلال در سیگنال است. برای اندازه گیری خوب، پروب باید به درستی قرار گیرد، اما حرکت بیش از حد جنین یا مادر باعث ایجاد نویز در سیگنال می شود. برای مقابله با این، یک پیش پردازش در پروژه تولدهای ایمن تر انجام شد تا مشخص شود کدام مناطق احتمالاً حاوی داده های نادرست هستند.

برخی از چالش‌های مجموعه داده این بود که سیگنال‌های FHR طول‌های مختلفی داشتند و برخی حاوی داده‌های واقعی بسیار کمی بودند. در این پایان نامه، سیگنال‌هایی که بیش از ۵۰ درصد از داده‌ها را از دست داده‌اند قابل استفاده در نظر گرفته نمی‌شوند و همراه با داده‌های فاقد برچسب‌های مربوطه از مجموعه داده مورد استفاده برای حل مشکل حذف می‌شوند.

مجموعه داده قابل استفاده نهایی شامل ۱۸۲۹ تولد است که ۱۶۶۵، ۹۱.۰۳٪ از آنها طبیعی بودند، و ۱۶۴، ۸.۹۷٪ بلافاصله پس از تولد نیاز به تهویه با کیسه ماسک داشتند. دلیل عدم تعادل این است که تمام داده‌ها از تولدهایی که کم‌خطر در نظر گرفته می‌شوند جمع‌آوری می‌شوند، که منجر به موارد کمتری می‌شود که نیاز به احیای فوری باشد. در شکل ۴.۱ و ۴.۲ می‌توانید نمونه‌هایی از سیگنال‌های قابل استفاده را مشاهده نمایید.



شکل ۱.۴: مثالی از سیگنالی که در آن BMV مورد نیاز نبود



شکل ۲.۴: مثالی از سیگنالی که در آن BMV مورد نیاز بود

۳.۴ افزایش داده ها

یکی از رایج ترین مشکلات در زمینه یادگیری ماشین کمبود داده یا کلاس های نامتعادل است. در این پایان نامه مجموع داده ها پس از پیش پردازش ۱۸۲۹ سیگنال بود که ۹۱ درصد آنها کلاس ۰ و ۹ درصد کلاس ۱ بودند. یکی از راه های مقابله با این مشکل افزایش داده است که به معنای افزایش تعداد سیگنال ها است. نقاط داده با توجه به حجم داده های موجود، کم نمونه گیری از بزرگترین طبقه منتفی شد و نمونه گیری بیش از حد از طبقه اقلیت به عنوان آزمایش انتخاب شد. تنها ۷۶۸۰ نمونه آخر در ابتدا برای طبقه بندی کننده استفاده شد که تنها بخشی از سیگنال های اصلی است. نمونه برداری بیش از حد شامل جابجایی این ۷۶۸۰ نمونه ۳ دقیقه به عقب بود تا زمانی که برای هر تولد ۱۰ تصویر طیف نگاری ایجاد شود.

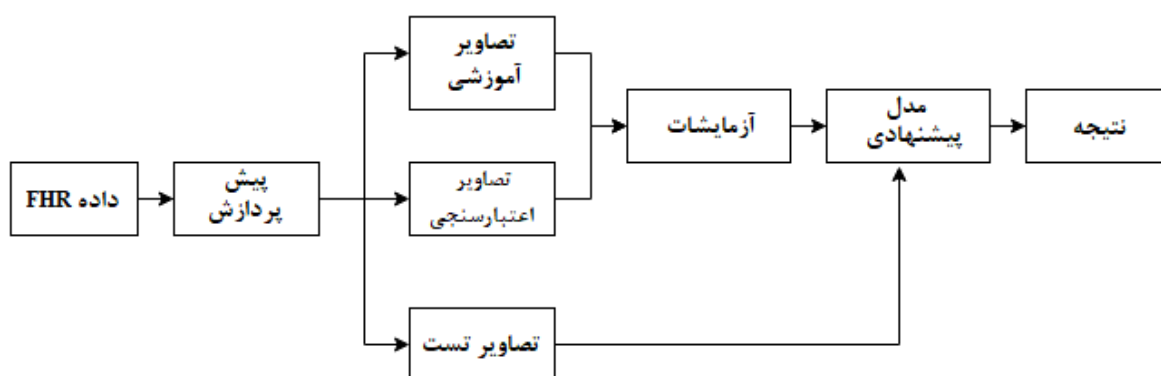
فصل پنجم

پردازش و روش های آن

فصل ۵

روش پیشنهادی

در این فصل هر بخش از سیستم پیشنهادی ارائه شده است. ابتدا پیش پردازش سیگنال های ورودی توضیح داده شده است. سپس شبکه های عصبی مورد استفاده و آزمایشات روی سیستم پیشنهادی توضیح داده شده است. نمای کلی سیستم در شکل ۵.۱ نمایش داده شده است.



شکل ۱.۵: نمای کلی سیستم پیشنهادی

۱.۵ پیش پردازش

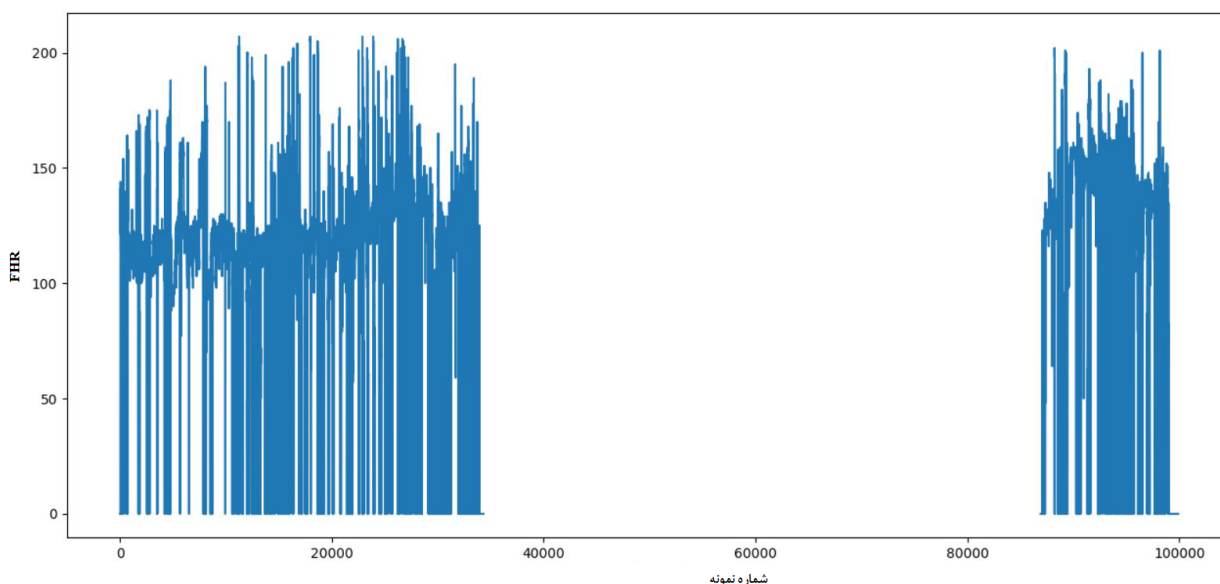
این بخش اجرای روش های پیش پردازش معرفی شده در بخش ۳ را توضیح می دهد که در شکل ۵.۳ نمایش داده شده است.

سیگنال های FHR مورد استفاده در این مطالعه دارای طول و کیفیت متفاوتی هستند. برای اینکه مجموعه داده قابل استفاده باشد، همه داده ها باید مراحل پیش پردازش مشابهی را طی کنند. بخش اول پیش پردازش، حذف داده هایی بود که برای استفاده بیش از حد نویز در نظر گرفته می شدند، به این معنی که سیگنال هایی حاوی کمتر از ۵۰ درصد داده های واقعی بودند، بدون احتساب زمانی که دستگاه مویو خاموش بود.

ماسک های نویز برجسب گذاری شده در سیگنال حذف شدند تا طبقه بندی کننده ویژگی های ایجاد شده توسط چیزی غیر از FHR را یاد نگیرد. همانطور که در تصویر دوم روی شکل نشان داده شده است، سیگنال با تعداد

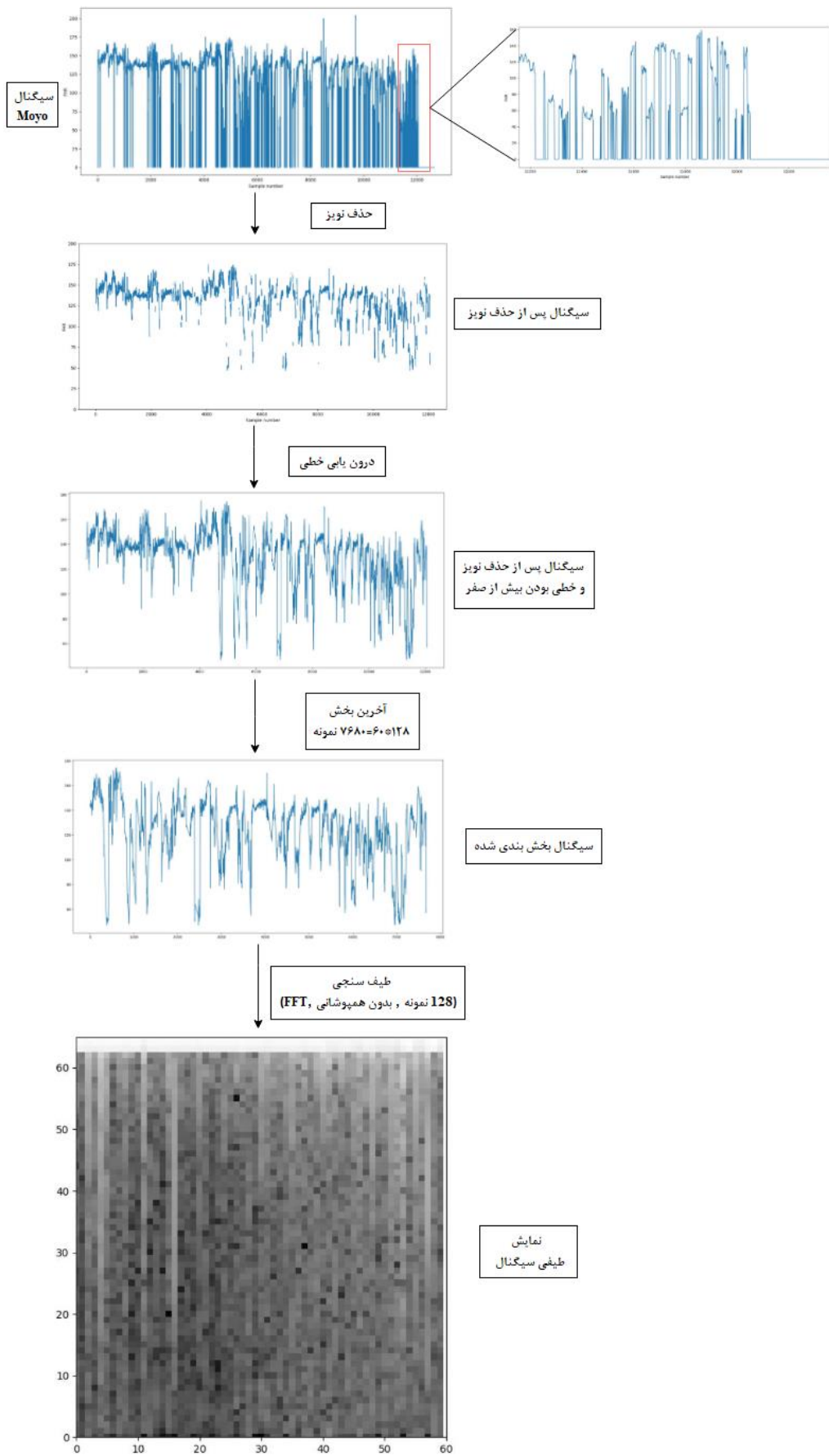
زیادی نمونه خالی باقی ماند. اینها به صورت خطی درون یابی شدند تا نمونه‌ها در محدوده ضربان قلب واقعی بیشتر شوند. نتیجه این کار در تصویر سوم قابل مشاهده است.

همانطور که در شکل ۵.۲ مشاهده می‌شود، شکاف‌های بزرگی در بیشتر سیگنال‌ها وجود داشت که در اثر خاموش شدن دستگاه مویو ایجاد می‌شد. با در نظر گرفتن این موضوع، طول‌های متفاوت سیگنال‌ها و این نظریه که بیشتر اطلاعات در قسمت آخر سیگنال قرار دارد، از هر یک تنها از ۷۶۸۰ نمونه آخر استفاده شد که کمی بیشتر از یک ساعت می‌باشد.



شکل ۲.۵: سیگنال نمونه

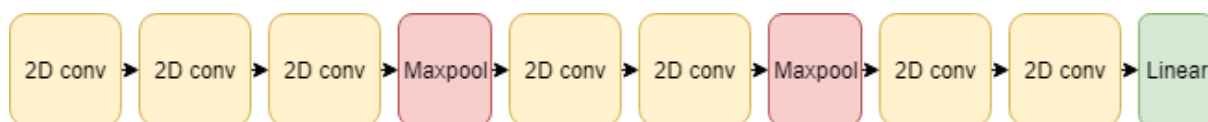
برای افزایش متمایز بودن ویژگی‌ها در سیگنال‌ها و استفاده از روش سنتی‌تر برای طبقه‌بندی، سیگنال‌ها با استفاده از ۶۰، ۱۲۸ قطعه طولانی و غیر همپوشانی، به نمایش‌های طیف‌گرام ۶۵ پیکسل در ۶۰ پیکسل تبدیل شدند. نتیجه این کار را می‌توان در تصویر پایین شکل ۵.۳ پیش پردازش مشاهده نمود.



شکل ۳.۵: نمای کلی پیش پردازش

۱.۱.۵ شبکه عصبی

چندین شبکه از جمله LSTM، RNN و CNN های مختلف مورد آزمایش قرار گرفتند. دو NN که بهترین نتایج اولیه را ارائه می‌دهند، یکی شامل ۷ لایه CL، ۲ لایه حداکثر و یک لایه خطی کاملاً متصل، با استفاده از نرمال‌سازی دسته‌ای، و ReLU به عنوان تابع فعال‌سازی بود. مدل دیگر یک مدل VGG-16 از پیش آموزش‌دیده بود، که در آن طبقه‌بندی‌کننده بر روی داده‌های FHR آموزش داده شد و وزن‌های استخراج ویژگی منجمد شدند. از دست دادن با استفاده از تلفات آنتروپی متقاطع محاسبه شد، زیرا به نظر می‌رسید که بهترین نتایج را داشته باشد.

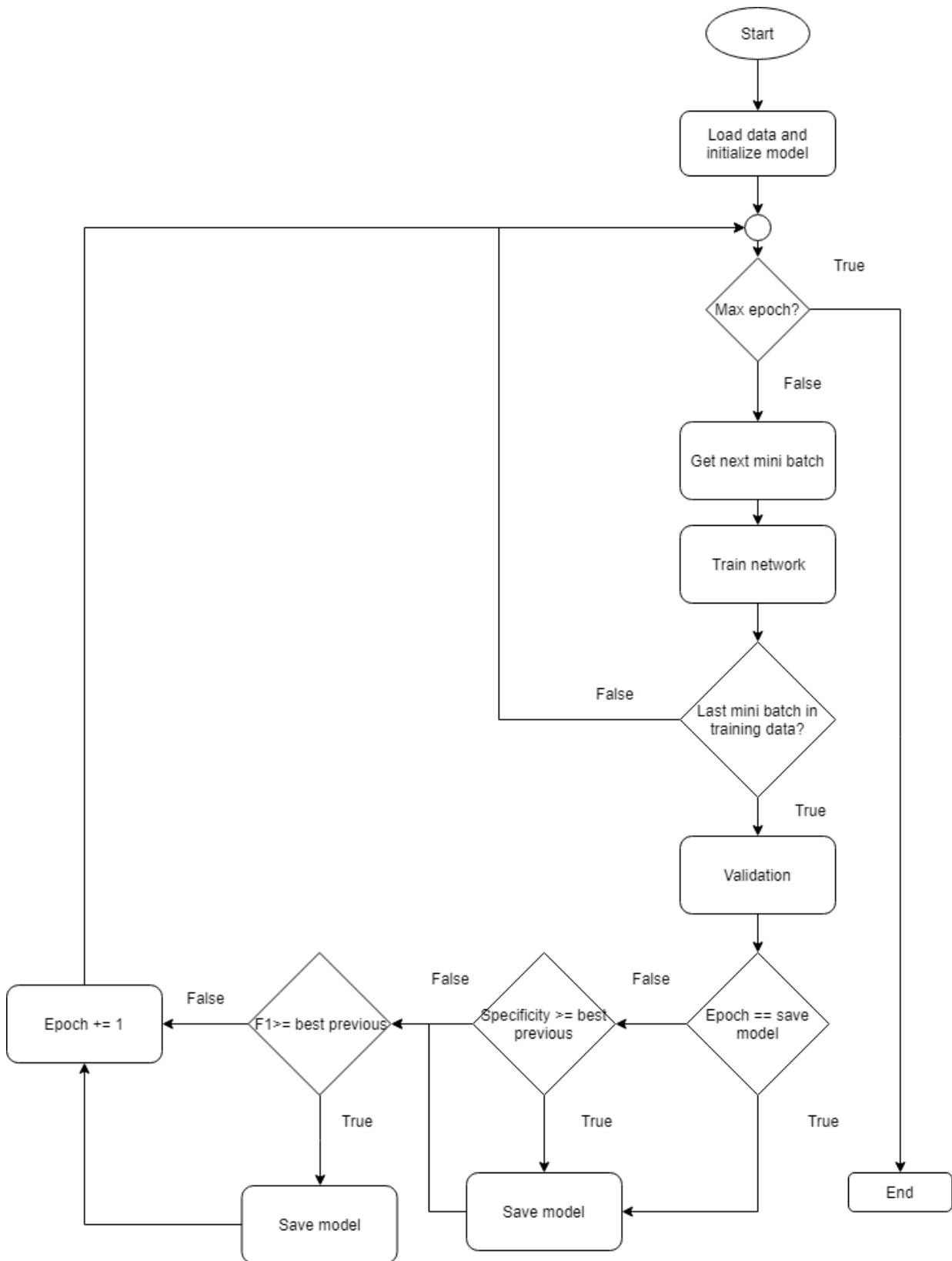


شکل ۴.۵: شبکه عصبی کانولوشن با ۷ لایه کانولوشن و ۱ لایه کاملاً متصل

منطق آموزشی برای NN در شکل ۵.۵ نشان داده شده است. مجموعه داده و مدل در حافظه بارگذاری شده و مقارنه اولیه می‌شوند. بعد از هر مینی بیچ، وزن‌ها به روز می‌شوند و اگر آخرین دسته آن دوره بود، از اعتبار سنجی عبور می‌کند. در مرحله اعتبارسنجی، عملکرد بر اساس داده‌های دیده نشده ارزیابی می‌شود. اگر دوره‌های کافی وجود داشته باشد، اگر ویژگی بهبود یابد یا اگر امتیاز F1 نسبت به بهترین مجموعه پارامترهای قبلی بهبود یابد، پارامترهای مدل ذخیره می‌شوند. بررسی بیش‌تر تمرینی به صورت دستی و توقف تمرین با تعیین حداکثر زمان انجام شد. تنها بهترین عملکرد مجموعه پارامترها برای هر مجموعه داده در مرحله آزمایشی پایان نامه استفاده شد.

۲.۵ آموزش انتقال

با توجه به حجم کم داده‌های موجود، آموزش انتقالی اجرا شد. استفاده از وزن یک شبکه از پیش آموزش‌دیده و انتقال بخشی از دانش از حوزه‌های قبلاً آموخته شده برای حل این کار. معماری و وزن از شبکه محبوب VGG-16 در این پایان نامه آزمایش شد.



شکل ۵.۵: فلوجارتی که فرآیند آموزش شبکه عصبی را تجسم می نماید [۵۳].

شبکه VGG-16 شامل ۱۳ لایه کانولوشن و ۳ لایه کاملاً متصل است. وزن‌های ۱۳ لایه کانولوشن، بخش استخراج ویژگی، منجمد شده‌اند، و فقط لایه‌های کاملاً متصل علامت‌گذاری شده، به نام طبقه‌بندی کننده، روی سیگنال‌های FHR آموزش داده می‌شوند. لایه خروجی نیز از پیش بینی ۱۰۰۰ کلاس به پیش بینی ۲ کلاس تغییر یافته است.

۱.۲.۵ تبدیل ورودی

شبکه VGG-16 به تصاویری با اندازه ۲۲۴ پیکسل در ۲۲۴ پیکسل به عنوان ورودی نیاز دارد. تصاویر طیف‌گرام اصلی ۶۵ پیکسل در ۶۰ پیکسل بودند و برای آزمایش یادگیری انتقال نیاز به تغییر اندازه داشتند. کتابخانه مشعل دارای چندین تبدیل داخلی است، از جمله تغییر اندازه تصاویر، که از یک فیلتر مثلث برای افزایش مقیاس استفاده می‌کند، و از این برای تغییر شکل تصاویر ورودی استفاده می‌شود.

۳.۵ آزمایش

مشکل طبقه بندی یک مشکل باینری بود. نیاز به BMV یا عدم نیاز به BMV. مجموعه داده به یک مجموعه آموزشی، ۶۰٪، یک مجموعه اعتبارسنجی، ۲۰٪، و یک مجموعه تست، به طور تصادفی پس از پیش پردازش اولیه داده‌ها، ۲۰٪ تقسیم شد. اینها برای همه شبکه‌های آزمایش شده یکسان بود. مجموعه اعتبارسنجی برای تصمیم‌گیری بهترین پارامترها و معماری برای مدل‌ها، آزمایش عملکرد مدل آموزش دیده‌شده بر روی داده‌های دیده نشده، استفاده شد. مجموعه آزمون برای آزمایش‌های نهایی استفاده شد که نتایج ارائه شده در این پایان نامه را ارائه می‌دهد. با توجه به عدم تعادل بین طبقات در مجموعه داده، دقت برای ارزیابی عملکرد رد شد. بنابراین از امتیاز F1 برای تصمیم‌گیری اینکه کدام معماری و ابرپارامترها بهترین عملکرد را برای هر آزمایش داشتند، استفاده شد.

پیش پردازش اولیه سیگنال‌های FHR برای همه آزمایش‌ها یکسان بود. حذف نویز برچسب‌گذاری شده، درون یابی خطی داده‌های از دست رفته در سیگنال، بخش‌بندی و تبدیل به تصاویر طیف‌گرام نرمال شده ۰-

۱ ۶۵ پیکسل در ۶۰ پیکسل. از همان مجموعه داده برای همه آزمایش‌ها استفاده شد. فرآپارامترهای بهینه در هر تغییر و مدل به صورت جداگانه جستجو شدند. از دست دادن با استفاده از تلفات آنتروپی متقاطع محاسبه شد، زیرا به نظر می‌رسید که در ابتدا بهترین نتایج را ارائه می‌دهد. تمامی شبکه‌های عصبی با استفاده از متلب ۲۰۲۲ و پایتون ۳.۷ شبیه‌سازی شده‌اند.

۱.۳.۵ آزمایشات قبلی

Run و LSTM-NN با تعداد لایه‌ها، نرخ یادگیری و ابعاد لایه‌های مختلف مورد آزمایش قرار گرفتند. این آزمایش‌ها هیچ نتیجه واقعی نداشت و به سرعت مشخص شد که تصاویر طیف‌نگاری و CNN عملکرد بهتری داشتند و این ایده‌ها کنار گذاشته شدند

۲.۳.۵ آزمایش ۱ - پایه

در آزمایش ۱، تصاویر طیف‌نگاری مستقیماً بدون پردازش بیشتر قبل یا بعد استفاده شد. در مجموع ۱۸۲۹ سیگنال قابل استفاده وجود داشت، ۱۶۶۵ مورد از آنها کلاس ۰، بدون نیاز به تهویه کیسه ماسک، و ۱۶۴ سیگنال کلاس ۱، نیاز به تهویه کیسه ماسک داشتند. برای مقابله با عدم تعادل داده‌های آموزشی، از یک تابع کاهش وزنی در طول آموزش مدل استفاده شد. وزن طبقه بندی اشتباه سیگنال متعلق به کلاس ۱ به شدت به اندازه ۱۱ طبقه بندی اشتباه سیگنال کلاس ۰ بهترین نتایج را در مجموعه اعتبار سنجی ارائه داد. در مجموع ۱۱۶۳ سیگنال آموزشی وجود داشت که ۱۰۶۳ سیگنال، ۹۱.۴٪ از آنها کلاس ۰ و ۱۰۰، ۸.۶٪ کلاس ۱ بودند. نتایج در فصل ۵ قابل مشاهده است.

۳.۳.۵ آزمایش ۲ - داده‌های افزوده شده

در آزمایش ۲، سیگنال‌های کلاس ۱ برای متعادل کردن کلاس‌ها تقویت شدند. فرآیند تقویت شامل استفاده از سیگنال‌های یکسان با جابجایی‌های مختلف در زمان بود. هر شیفت ۳۶۰ نمونه، مربوط به ۳ دقیقه بود.

این ۹ بار برای هر تولد در کلاس ۱ انجام شد، به این معنی که حداقل طول سیگنال ۱۰۹۲۰ نمونه مورد نیاز بود. تعداد کل تصاویر در مجموعه آموزشی اکنون ۱۸۶۹ بود که ۹۵۹ در کلاس ۰ و ۹۱۰ در کلاس ۱ توزیع شده است. تعداد کل سیگنال ها در کلاس ۱ به طور چشمگیری افزایش یافته است، اما تعداد بیماران از ۱۰۰ به ۹۱ کاهش یافته است.

۴.۳.۵ آزمایش ۳ - پیش و پس پردازش

آزمایش سوم از همان مدل استفاده کرد که برای آزمایش ۲ آموزش داده شد. تفاوت در پیش پردازش و پس پردازش است. به جای اینکه فقط کلاس ۱ تقویت شود، همه سیگنال ها برای ایجاد ۱۰ تصویر طیف گرا برای هر بیمار تقویت شدند. مجموعه اعتبارسنجی آزمایش ۳ شامل ۲۸۰۰ تصویر است که ۲۵۰۰ تصویر، ۸۹.۳ درصد کلاس ۰ و ۳۰۰، ۱۰.۷ درصد کلاس ۱ است.

تمام تصاویر طیف‌نگاری ابتدا به صورت جداگانه طبقه‌بندی شدند، مشابه آنچه در آزمایش‌های قبلی انجام شد، سپس تمام طبقه‌بندی‌ها برای یک تولد ترکیب و آستانه‌گذاری شدند. آستانه ارائه بهترین عملکرد در مجموعه اعتبار سنجی انتخاب شد.

۵.۳.۵ آزمایش ۴ - انتقال یادگیری

آخرین آزمایش، تکرار تمام آزمایش‌های قبلی بود، اما به جای استفاده از وزن‌های اولیه تصادفی، از مدل VGG-16 از پیش آموزش‌دیده برای مقداردهی اولیه مدل استفاده شد. بخش تشخیص ویژگی مدل منجمد شد و تنها طبقه‌بندی‌کننده با استفاده از داده‌های FHR آموزش داده شد. لایه خروجی به تناسب یک مشکل کلاس ۲ تغییر یافت.

فصل ششم

نتایج و تجزیه و تحلیل

فصل ۶

نتایج

این فصل شامل ارزیابی آزمایش های ارائه شده در بخش ۴.۳ است. میزان یادگیری از ۰.۰۰۰۱ تا ۰.۰۰۵، افت تحصیلی بین ۰ تا ۰.۲ و وزن های مختلف مورد آزمایش قرار گرفت. نتایج ارائه شده از بهترین مجموعه برای هر یک از آزمایش ها استفاده می کند.

۱.۶ نتایج و تجزیه و تحلیل

همه مدل ها بر روی مجموعه آموزشی آموزش داده شدند و فرآیندها بر اساس عملکرد در مجموعه اعتبارسنجی انتخاب شدند. بهترین امتیاز F1 هر مدل برای هر مجموعه داده در جدول ۱.۶ ارائه شده است.

جدول ۶.۱: امتیاز F1 در طول اعتبارسنجی. * آستانه ۰.۲ برای مدل ۸ لایه و ۰.۵ برای مدل از پیش آموزش دیده

آزمایش شماره ۴ ۸ لایه ایجاد شده

آزمایش شماره ۱	۰.۲۳۶۸	۰.۲۱۳۵
آزمایش شماره ۲	۰.۲۷۵۲	۰.۲۷۷۲
آزمایش شماره ۳*	۰.۲۵۳۲	۰.۲۱۱۱

بهترین روش عملکرد برای مجموعه اعتبارسنجی، مدل ۸ لایه ای بود که بر روی داده های آموزشی متعادل تر با داده های کلاس ۱ افزوده شده برای ارائه ۱۰ تصویر در هر تولد آموزش داده شد. پارامترهای بیشتر در جدول صفحه بعد نشان داده است:

جدول ۶.۲: پارامترهای مدل مورد استفاده در آزمایش ۳

پارامتر	مقدار
مجموعه قطار	۱۸۶۹ تصویر
مجموعه اعتبارسنجی	۳۳۳ تصویر
دوره ها	۳۳
میزان یادگیری	۰.۰۰۳
تابع فعال سازی	ReLU
ترک تحصیل	۰.۱
اندازه کوچک دسته ای	۳۳۳

همانطور که جدول ۶.۳ نشان می دهد، شبکه تشخیص داده است که ۴۳.۷۵ درصد از تولدها مورد نیاز BMV بوده است، با دقت کلی ۷۸.۰۸ درصد.

VGG-16 از پیش آموزش دیده نیز بهترین عملکرد را در داده های تمرینی متعادل داشت و تقریباً همان امتیاز f1 را به دست آورد. پارامترها در جدول زیر نشان داده شده است:

جدول ۶.۳: ماتریس سردرگمی اعتبارسنجی در مدل ۸ لایه

		پیش بینی شده	
		C ₀	C ₁
واقعی	C ₀	۲۴۶	۵۵
	C ₁	۱۸	۱۴

جدول ۶.۴: عنوان

پارامتر	مقدار
مجموعه قطار	۱۸۶۹ تصویر
مجموعه اعتبارسنجی	۳۳۳ تصویر
دوره ها	۳
میزان یادگیری	۰.۰۰۲۳
اندازه کوچک دسته ای	۱۱۱

جدول ۶.۵: ماتریس سردرگمی اعتبارسنجی در مدل VGG-16 از پیش آموزش دیده

		پیش‌بینی شده	
		C ₀	C ₁
واقعی	C ₀	۲۳۹	۶۲
	C ₁	۱۷	۱۵

اندازه مینی بیچ به دلیل محدودیت حافظه در GPU موجود باید کاهش یابد. مدل های از قبل آموزش دیده برای رسیدن به بهترین نتیجه به زمان کمتری نیاز داشتند و از نرخ یادگیری پایین تری استفاده کردند. این مدل ۴۶.۸۷۶ درصد از عوارض زایمان را با دقت ۷۵.۶۸ درصد تشخیص داد.

۲.۶ تأیید نتیجه

بهترین عملکرد شبکه از اعتبارسنجی بیشتر مورد ارزیابی قرار گرفت. در آزمایش‌های قبلی، پارامترهای مدل با توجه به عملکرد آن بر روی داده‌های اعتبارسنجی انتخاب شدند. برای تایید نتایج، مدل بر روی مجموعه تست آزمایش شد. تنها بهترین مدل با عملکرد در مجموعه اعتبارسنجی به این روش آزمایش شد. نتایج تایید در شکل های زیر نشان داده شده است.

جدول ۶.۶: ماتریس سردرگمی آزمون نهایی. امتیاز F1 برای کلاس ۱ برابر با ۰.۱۳۲۰۷ می‌باشد.

پیش‌بینی شده

		C ₀	C ₁
واقعی	C ₀	۲۳۹	۶۲
	C ₁	۱۷	۱۵