FACULTY OF ENGINEERING

DEPARTMENT OF ELECTRICAL, COMPUTER AND BIOMEDICAL ENGINEERING

MASTER'S DEGREE IN BIOENGINEERING

MASTER THESIS

AI-Based Prediction of Fetal Outcomes in Congenital CMV from Fetal Brain MRI: A Feasibility Study

Predizione degli outcome fetali nel CMV congenito tramite l'applicazione dell'intelligenza artificiale alla RM cerebrale fetale: uno studio di fattibilità

Supervisor: Dr. Enea Parimbelli

Candidate: Elisa Pitzalis

A.Y. 2024/2025

To my family, for their endless support

Abstract

Cytomegalovirus (CMV) is one of the most common pathogens worldwide, with a prevalence of 83% in the general population [1]. Although infections are usually asymptomatic in healthy individuals, congenital cytomegalovirus (cCMV) represents the leading infectious agent for birth defects [2] and the primary cause of nongenetic sensorineural hearing loss (SNHL) [3],[4].

While magnetic resonance imaging (MRI) provides detailed evaluation of fetal brain abnormalities, early detection remains a major challenge, due to incomplete fetal brain development and the delay between infection and lesion manifestation [3]. In this context, predictive modeling has gained increasing relevance in medical research [46], yet its application to fetal imaging in cCMV remains unexplored.

This thesis presents as an explorative study aimed at evaluating whether MRI data acquired at 20 weeks of gestation contains sufficient information to predict corresponding findings at 30 weeks. To address this, adapted *Vision Transformer Masked Autoencoder* (ViT-MAE) models were employed for both anomaly detection and classification tasks.

The results show that the models are able to capture global relationships in the data and effectively reconstruct images with reasonable accuracy; however, they struggle to reconstruct finer-grained details and to reliably distinguish the two outcomes in both tasks. This limitation was primarily influenced by the lack of brain segmentation, which biases the reconstruction error on with ViT-MAE models rely, thereby affecting the assessment of diagnostic potential in early gestation data,

Overall, this pilot study establishes a methodological foundation for future research, highlighting key challenges such as limited data availability, while also providing clear directions for future work.

Abstract (in italiano)

Con una prevalenza dell'83% nella popolazione generale, il citomegalovirus rappresenta uno dei patogeni più comuni al mondo [1]. Nonostante negli individui sani le infezioni siano generalmente asintomatiche, l'infezione congenita da citomegalovirus (cCMV) costituisce la principale causa infettiva di difetti alla nascita [2] e di ipoacusia neurosensoriale (SNHL) non genetica [3],[4].

In questo ambito, la risonanza magnetica (RM) consente una valutazione dettagliata delle anomalie cerebrali fetali, ma la diagnosi precoce di queste ultime costituisce ancora una sfida significativa, a causa dello sviluppo incompleto del cervello fetale nelle prime fasi della gestazione e del tempo che intercorre tra l'infezione e la comparsa delle lesioni [3].

Sebbene in campo diagnostico i modelli predittivi abbiano acquisito una crescente rilevanza [46], la loro applicazione all'imaging fetale nel cCMV rimane ancora inesplorata.

Per le suddette ragioni, il presente lavoro di tesi consiste in uno studio esplorativo volto a valutare se i dati di risonanza magnetica acquisiti alla 20ª settimana gestazionale contengano informazioni sufficienti per predire i corrispondenti esiti alla 30ª settimana. A tal fine, sono stati impiegati i cosiddetti *Vision Transformer Masked Autoencoder* (ViT-MAE), adattati opportunamente al fine di svolgere sia un task di rilevazione di anomalie, che un compito di classificazione.

I risultati dello studio mostrano che i modelli sono in grado di imparare relazioni globali dai dati e ricostruire le immagini con discreta accuratezza; tuttavia, essi presentano difficoltà nel ricostruire i dettagli più fini, e quindi nel distinguere in maniera affidabile i due possibili esiti diagnostici, quali patologico o negativo. Nel presente lavoro, questa limitazione è stata principalmente attribuita all'assenza di segmentazione cerebrale, in quanto il background introduce un bias nell'errore di ricostruzione su cui si basano i modelli ViT-MAE. Questo inficia perciò la valutazione del potenziale diagnostico suddetto nelle fasi precoci della gestazione.

Nel complesso, questo studio di fattibilità stabilisce le basi metodologiche per ricerche successive ed evidenzia eventuali sfide chiave, come la limitata disponibilità di dati, fornendo al contempo indicazioni chiare per sviluppi futuri.

Summary

Abstract **5**

Abstract (in italiano) **6**

1. Introduction **10**

1.1. Clinical background *10*

1.1.1. Human cytomegalovirus 10

1.1.2. Congenital cytomegalovirus 11

1.1.3. Clinical management of cCMV 16

1.1.4. Fetal MRI in cCMV 19

1.2. Potential role of AI *26*

1.3. Objective of the thesis *28*

1.4. Related work *28*

2. Materials and methods **30**

2.1. Study design *30*

2.2. Dataset description *31*

2.2.1. Study population 31

2.2.2. Data exploration 31

2.3. Preprocessing *32*

2.3.1. MRI data selection: sequences and planes 32

2.3.2. Slice trimming and data standardization 34

2.3.3. Outcome definition 36

2.4. AI models *39*

2.4.1. Original Vit-MAE architecture 39

2.4.2. Proposed framework 41

3. Results **44**

3.1. Anomaly detection *44*
3.1.1. Best model selection: training and validation losses 44
3.1.2. Comparison of reconstructions with different patch sizes 47
3.1.3. Reconstruction error visualization 48
3.1.4. Quantitative assessment of reconstruction performance 52
3.2. Classification task *53*
3.2.1. Selection of the best pretrained model 53
3.2.2. PCA and UMAP analysis 56
3.2.3. Finetuning results 59
4. Discussion and conclusions 60
4.1. Findings' interpretation *60*
4.1.1. Anomaly detection 60
4.1.2. Classification task 62
4.2. Limitations of this study *64*
4.3. Discussion of research objectives *65*
4.4. Future work *66*
Acknowledgements 69
References 71

1. Introduction

Congenital cytomegalovirus (cCMV) infection represents a significant cause of neurodevelopmental impairment, making early detection and precise evaluation essential in clinical practice [3]. In this context, artificial intelligence (AI) may provide valuable support in identifying patterns that are not easily detectable with visual inspection alone.

This chapter provides an overview of the clinical background, including symptoms, epidemiology and clinical management of the infection. Subsequently, it discusses the potential role of artificial intelligence and presents the objective of the thesis, along with a review of the state of the art.

1.1. Clinical background

1.1.1. Human cytomegalovirus

Human cytomegalovirus (HCMV) is a ubiquitous β-herpesvirus characterized by high genetic complexity, presenting the largest genome amongst viruses that infect humans. This reflects the virus's refined strategies to entry and attack its host [5], as well as those to inhibit the host to defend themselves [6].

Transmission happens via direct contact of infectious body fluids, such as saliva, urine, blood, etc. The disease goes generally unnoticed in immunocompetent individuals, who might manifest mild flu-like symptoms – e.g. fever, headache, loss of appetite, fatigue, swollen glands – or no symptoms at all [3]. About 10% of infected older children and adults also develop a heterophile-negative mononucleosis syndrome, which represents the CMV-counterpart to the more common Epstein-Barr virus-induced mononucleosis [7].

Symptoms are decreasingly noticeable or even absent with pre-existing immunity, i.e. during a secondary infection [3]. In fact, following primary infection, HCMV establishes **latency**, i.e. a state in which the virus remains silent in the host, but can periodically reactivate and cause a secondary infection, showing a typical pattern amongst viruses of the

Herpesviridae family, that allows the host to spread the virus again [5]. This behavior is attributable to specific genes that interfere with the host's natural killer cells, preventing viral clearance and enabling long-term infection and periodic reactivation [6].

Correspondingly, the host develops partial immunity towards the virus, thanks to the generation of a robust population of CMV-specific antibodies and effector-memory T-cells, that are maintained over time. This adaptive mechanism, known as *memory inflation*, does not eliminate the virus, but keeps it under control and reduces symptoms in upcoming reactivations [8].

For these reasons, HCMV does not show seasonality and is highly prevalent in the general population, with an estimated global seroprevalence of 83% (95% CI, 78–88). The prevalence is 86% among women of reproductive age (95% CI, 83–89), and 86% among blood or organ donors (95% CI, 82–89) [9].

Specific populations show a higher incidence of contracting the disease, such as nursing infants, toddlers, sexually active teenagers, as well as those who work in group daycare settings (e.g. teachers, school staff). Individuals who have regular contact with children are more vulnerable, as children under two years of age may shed the virus in their urine and saliva for up to 2 years [10],[7]. Thereby, pregnant mothers with at least one other child are particularly at risk [11], as their infection could lead to intrauterine transmission, which will be discussed in the following paragraph.

1.1.2. Congenital cytomegalovirus

It has been estimated that between 1% and 4% of women experience a primary CMV infection during pregnancy. This event is commonly termed **maternal seroconversion** or **maternal primary infection (MPI),** indicating the absence of preconceptional immunity. Additionally, CMV might reactivate during pregnancy, or the mother might catch a different viral strain, leading to **non-primary infections (NPIs**), which happens in about 10-30% of women with pre-existing immunity. The hormonal and immunological changes that occur throughout pregnancy may facilitate this event [7].

However, maternal infection does not necessarily imply fetal infection, nor fetal or neonatal severe symptomatology. In fact, current evidence suggests that 0.6-6.1% of newborns are affected by CMV every year [10].

As reported by Chatzakis et al. [12], the overall rates of transmission, fetal damage and presence of symptoms at birth are summarized in Table 1, and will be further analyzed in the following paragraphs. These rates are based on a total of 2942 fetuses across 10 studies, and are stratified according to the timing of maternal infection.

Table 1. Rates of transmission, neonatal infection and presence of symptoms at birth [12].

	Vertical transmission rate	**Fetal damage if fetus is infected**	**Symptomatic at birth**
Pre-conceptional	5.5% (95%, CI: 0.1-10.8%)	NA	NA
Periconceptional	21% (95%, CI: 8.4-33.6)	28.8% (95% CI: 2.4-55.1	1.3% (95% CI: 0-4.5)
First trimester	36.8% (95% CI: 31.9-41.6)	19.3% (95% CI: 12.2 – 26.4)	9.1% (95% CI: 2.7-15.6)
Second trimester	40.3% (95% CI: 35.5-45.1)	0.9% (95% CI: 0-2.4)	0.3% (95% CI: 0-1.1)
Third trimester	66.2% (95% CI: 58.2-74.1)	0.4% (95% CI: 0-1.5)	0.4% (95% CI: 0-1.6)

Transmission to the fetus

In both MPIs and NPIs the virus can cross the placenta and infect the fetus; however, the likelihood of transmission is different, as well as the sequelae, i.e. the lasting effects after the acute phase of the illness.

More specifically, MPIs have a 30-40% risk of vertical transmission, i.e. mother-to-fetus infection, with rates increasing from 20-30% in early pregnancy to 40-70% when the mother

is infected during the third trimester [13]. In Europe, where CMV seroprevalence among pregnant women ranges from 50 to 85%, about half of cCMV cases result from MPIs [14].

Additionally, a 2019 study conducted in France found that, for MPIs, the interval between subsequent pregnancies plays a key role. Specifically, women who become seropositive during a second pregnancy, within a 2-year span from their previous pregnancy, have the highest risk of vertical transmission – 88% of all MPIs, and 92% of first trimester MPIs– as well as the greatest risk of sequelae, respectively 24- and 6-fold higher than general pregnant population [15].

On the other hand, only 1-3% of NPIs appear to result in intrauterine infections [7]. However, it should be noted that transmission rates – especially by trimester – are highly uncertain for NPIs, since diagnoses for secondary infection are challenging [13]. Because the patient may become infected at any time, multiple serological tests over time would be needed to verify seronegativity; however, this is clearly not feasible, leaving a gap in epidemiological data. Additionally, previous studies have focused mainly on MPIs, as they were considered the only source of adverse sequelae [16], resulting in even less data on NPIs.

Vertical transmission can also occur intrapartum or postnatally, through exposure to infected cervical secretions during delivery or ingestion of infected breast milk [7].

Symptoms and sequelae

Being still immunologically immature, the fetus is particularly vulnerable, especially if the mother contracts the virus in the periconceptional period or in the first trimester of pregnancy, due to the high rate of cell replication [17],[3].

Specifically, CMV can infect cytotrophoblasts in the placenta, making it dysfunctional and facilitating vertical transmission, to subsequently target neural stem cells – progenitors of neurons and glia cells – which leads to reduced brain volume and abnormal neuronal migration [12], [18]. Additionally, CMV influences the host's inflammatory and immune

response, resulting in brain injuries, fetal growth restriction, preterm birth or even miscarriage [1],[11],[19].

However, in some cases, the damage is not clinically apparent at birth. In newborns, CMV can be either symptomatic or asymptomatic – the former being less frequent, involving only 10-20% of babies [3] – but it can also manifest later in life, by generating sequelae.

In symptomatic cases, common manifestations include visible signs, laboratory abnormalities and/or intracranial abnormalities, such as microcephaly, intrauterine growth restriction (IUGR) and hepatosplenomegaly. Additionally, 40-58% of symptomatic newborns will also develop long-term consequences in the future, such as vision deficits, mental or developmental delay, seizures and cerebral palsy [7]. For instance, SNHL may manifest at birth or progress over the course of childhood, affecting 15-25% of all children with congenital CMV [6].

Neurological damage is primarily associated to imaging findings, which will be discusses in the following section. The complete list of possible symptoms and clinical findings in symptomatic newborns is presented in Table 1, as of the 2015 conference of the European Society of Pediatric Infectious Diseases [20].

Table 2. Possible symptoms and clinical findings caused by cCMV in symptomatic newborns, as of the 2015 conference of the European Society of Paediatric Diseases [20].

CLINICALLY DETECTABLE SYMPTOMS/SIGNS
Physical examination Small for gestational age (birth weight < −2 standard deviations for gestational age) Microcephaly (head circumference < −2 standard deviations for gestational age) Petechiae or purpura (usually found within hours of birth and persist for several weeks) Blueberry muffin rash (intra dermal hematopoiesis) Jaundice (can be present at the first day after birth and usually persists longer than physiologic jaundice) Hepatomegaly Splenomegaly
Neurologic physical examination Microcephaly (head circumference < −2 standard deviations for gestational age) Neurologic signs (lethargy, hypotonia, seizures, poor sucking reflex)
ABNORMALITIES DETECTED INCIDENTALLY OR THROUGH SUBSEQUENT INVESTIGATION/SPECIALIST EXAMINATION

Laboratory results Anemia Thrombocytopenia (occurs in the first week but platelets often increase spontaneously after the second week) Leukopenia, isolated neutropenia Elevated liver enzymes (alanine aminotransferase/aspartate aminotransferase) Conjugated hyperbilirubinemia
Cerebrospinal fluid Abnormal cerebral fluid indices, positive CMV DNA
Neuroimaging Calcifications, periventricular cysts, ventricular dilatation, subependymal pseudocysts, germinolytic cysts, white matter abnormalities, cortical atrophy, migration disorders, cerebellar hypoplasia, lenticulostriatal vasculopathy
Hearing test Sensorineural hearing loss uni- or bilaterally
Visual examination Chorioretinitis, retinal hemorrhage, optic atrophy, strabismus, cataracts

Asymptomatic infections can lead to fetal brain injuries and long-term sequelae as well, generally milder than those observed in symptomatic infections at birth [3],[7]; however limited data are available on late-onset symptoms, due to often incomplete follow-up [21].

Overall, infections in early pregnancy and MPIs are associated with worse outcomes, especially in terms of severe hearing impairment, showing an inverse trend with viral transmission [3], [12]. However, also the late gestational phase presents a risk for sequelae, particularly for symptomatic children at birth.

On the other hand, the difference between MPIs and NPIs in terms of long-term consequences remains unclear. Earlier literature focused exclusively on MPIs, due to the mistaken belief that maternal seroimmunity protected against cCMV [16]; nowadays, it is widely recognized that both MPIs and NPIs can contribute to adverse outcomes [13], [22]. However, it is still unclear whether they are responsible for different degrees of risk, with multiple studies showing opposite results [13]-[15]-[4], [23] [13], [24].

CMV infection may also be contracted after birth. However, intrapartum or postnatal infections rarely cause symptoms or sequelae; some studies reported clinical deterioration in infected pre-term infants (e.g. worsening respiratory status, neutropenia and septic-like manifestations), while others have not found an association [7].

1.1.3. Clinical management of cCMV

Given its potential consequences, early diagnosis of cCMV is essential to offer better treatment options, and, in severe cases, allow consideration of pregnancy termination. However, due to its complex clinical management, significant gaps in knowledge persist, and extensive research efforts are still ongoing.

In this context, the 2024 European guidelines for the prenatal, neonatal and postnatal management of congenital CMV, developed under the patronage of the European Society of Clinical Virology, have provided the most comprehensive and up-to-date framework for the management of cCMV. The resulting recommendations include strategies to prevent vertical transmission, available screening, diagnosis and treatment options, as well as prognosis and long-term follow-up [25].

The following sections outline these aspects, with particular attention to the distinctions between maternal, fetal and neonatal management.

Maternal screening

Even if CMV meets the established criteria for population screening – being clinically significant, well-defined and prevalent – routine universal screening, either antenatal or neonatal, is still not recommended. The main reasons are that, as of now, it is still difficult to prevent transmission and to predict sequelae [26].

Neonatal screening will be discussed in the following section. As for maternal infections, CMV screening is implemented only in some countries [25]. Multiple studies demonstrated that universal screening and valaciclovir therapy – a common medication to prevent mother-to-fetus transmission – would be more cost-effective than current practice [3], [27]. In fact,

early pregnancy screening could identify seronegative women, who could then be targeted with specific preventive measures [26]. For this reason, the European guidelines recommend to test mothers for CMV in the first trimester of pregnancy, and repeat it every 4 weeks until 14-16 weeks in confirmed or suspected seronegative women, testing beyond that only in symptomatic mothers [25], [26].

Maternal screening can be performed through serological testing with CMV-specific immunoglobin G (IgG) and M (IgM) antibodies, which indicate respectively a past infection and a recent/ongoing infection. Specifically, latest generation CMV IgG have high sensitivity (97-100%) and specificity (96-100%), and IgG avidity can also help to estimate the timing of the infection, potentially excluding recent MPIs [25]. CMV could also be detected through PCR in blood and urine, but it is not recommended for suspected NPIs nor for timing purposes, and only helpful after Ig testing to exclude or confirm an ongoing MPI [25].

NPIs could go undetected in this process [25], and these measures do not identify who has a higher likelihood of transmitting the virus, adding to mothers the stress of having extra tests, as well as an increased likelihood of miscarriage [26].

Diagnosis of fetal and neonatal infection

Once maternal infection has been established, the next step is to assess fetal and neonatal involvement.

Fetal diagnosis

The best diagnostic tool for fetal CMV infection is CMV PCR in the amniotic fluid, ideally performed at 17 weeks or beyond, provided that maternal seroconversion occurred at least 8 weeks earlier. In these conditions, CMV PCR has a sensitivity of 87-95% and specificity of almost 100% [25]. Some women, however, may refuse amniocentesis; in these cases, subsequent fetal ultrasounds are the recommended strategy [25].

Neonatal screening

While fetal diagnosis provides essential information to in-utero infection and potential prognosis, newborn screening remains crucial to confirm congenital CMV; however, only in 2024 guidelines have recommended it for all newborns delivered by infected mothers, as well as in those who present imaging abnormalities compatible with CMV [25].

Neonatal infection can be diagnosed by PCR testing of urine or saliva, obtained within the first three weeks of life – ideally right after birth – to differentiate congenital from postnatal infection [25]. Additionally, studies on PCR testing on dried blood spots (DBS) are ongoing, with an estimated sensitivity of roughly 80% [25], [26]. Neonatal testing through PCR for IgG and IgM is not recommended, since it cannot confirm infection [25].

In any case, establishing a neonatal differential diagnosis can be challenging, since nonspecific laboratory abnormalities are common and other congenital conditions (e.g. toxoplasmosis, syphilis, rubella, HSV) have similar nonspecific symptoms.

Prevention of fetal infection and neonatal treatment

Once maternal infection occurs, pharmacological treatment to prevent transmission or reduce sequalae involves hyperimmune globulin (HIG) and antiviral drugs, such as valaciclovir and valganciclovir, with valaciclovir currently considered the gold standard [3], [25]. Treatment efficacy appears to be time-dependent, with earlier initiation leading to greater effectiveness [25].

Similarly, newborns with symptomatic cCMV should be treated with antivirals as well for 6 months, with careful periodical monitoring given their potential toxicity [25], [28], [29].

Prognosis and long-term follow-up

Once congenital CMV infection is diagnosed, evaluation of organ involvement should be the next step. During pregnancy, a combination of ultrasound imaging and MRI is

recommended to predict the prognosis, presenting a high negative predictive value for moderate to severe sequelae. However, there is still a residual risk of hearing loss at birth, of around 17% [25]. Neuroimaging for cCMV will be further analysed in Section 1.1.4.

After birth, a physical examination should be performed, including ophthalmologic and audiological assessments, followed by laboratory and neurosonographic evaluations. This allows to distinguish between asymptomatic and symptomatic infants, which defines timing and extension of follow-up [3], [25]. In fact, follow-up could last up to 6 years of age, because of possible delayed SNHL and neurodevelopmental changes that emerge at school entry [25], [26].

1.1.4. Fetal MRI in cCMV

Prenatal ultrasound (US) represents the first-line fetal imaging modality to assess fetal development. However, when US results are inconclusive or irregular, magnetic resonance imaging (MRI) becomes the second best option, as it provides a more detailed evaluation of brain abnormalities in congenital conditions, while also having many advantages in terms of safety and spatial resolution [3], [25]. Additionally, it overcomes some US limitations, such as maternal obesity, fetal position and oligohydramnios [30].

MRI principles and sequences

The basis of MRI technology lies on the behaviour of hydrogen protons in the body, whose orientation can be modified through the utilization of an external magnetic field, a radiofrequency (RF) pulse and magnetic field gradients, which help generating a measurable signal. Two timing parameters are T_1 and T_2, whose values vary across human tissues; these differences enable MRI to distinguish between various types of tissue and create the contrast shown in MRI images [31].

Magnetic resonance sequences consist of specific combinations of different RF pulses and gradient manipulations. Depending on parameters selection, particularly the TR (repetition

time) and TE (echo time), different type of sequences can be obtained, which can result in predominantly T_1- or T_2-weighted images [32].

T_1-weighted (T1W) sequences are characterized by short TR and short TE. On T1W series, fat, subacute hematomas, slow-moving blood and gadolinium base MR contrast appear bright, while tissues containing fluid (such as water or cerebrospinal fluid) appear dark. Conversely, T2-weighted (T2W) sequences have both a long TR and long TE, making free fluid and tissues with a high free water content appear bright [32].

Thus, T1W images are especially useful for displaying anatomical and morphological details; when analyzed in conjunction with T2W images, they allow classifications of lesions as iso-, hypo- or hyper-intense in relation to grey matter, white matter, CSF or fat [32].

However, it is possible to distinguish two main categories of MRI pulse sequences: spin echo (SE) and gradient recalled echo (GRE). All other sequences are variations of these, with additional parameters [30]. Among the SE sequences, because of their shorter acquisition intervals, *fast spin echo* (FSE or TSE) have effectively supplanted traditional SE T2W sequences for imaging of brain and spine [32]. The main reasons are that they effectively reduced fetal motion artifacts and made fetal sedation unnecessary [30]. Fat appears rather bright, while blood looks dark; however these scans are much less sensitive to hemorrhages and lesions in the peripheral cortex or the periventricular region, due to the bright CSF signal; thus, they are usually employed with GRE sequences [32].

The **single-shot fast spin-echo** (**SS-FSE** or T2-SSH-TS) is the standard sequence in fetal imaging. Since SS-FSE relies on single slice acquisition, only the slice that is being collected is at risk of artifacts in case of fetal motion [30].

On the other hand, T1W images are usually obtained through 2D GRE sequences, in which every slice is captured at the same time. Therefore, any little movement at any point throughout the acquisition would deteriorate the series. Furthermore, spatial resolution is worse than SS-FSE [30].

Among the GRE sequences, two categories are incoherent (or spoiled) and coherent sequences (also known as steady-state free precession [SSFP]) [32]. Particularly, balanced steady-state-free precession (bSSFP) sequences provide a T2/T1 image contrast with a very high signal-to-noise ratio. This type of sequence uses **balanced gradient pulses** to keep the stimulated magnetization in the transverse plane. Because of the bright blood signal, this technique is especially suitable for the evaluation of the heart and vessels [33].

CMV imaging protocol

The two primary goals of imaging in cCMV are the detection of fetal structural anomalies for accurate diagnosis and provision of prognostic information [3].

While the 2024 ECCI consensus recommendation has confirmed the superiority of MRI in cCMV prognosis evaluation comparatively to US, it has not established a detailed MR imaging protocol. Therefore, just as in previous years, distinct facilities follow different practices centered on a variety of factors, including experience, sequence and scanner availability, and acquisition findings [3], [25].

According to Diogo et al. [3], whose review focuses specifically on MRI for cCMV, T2W images of fetal body and brain should be collected in three planes, while T1W and DWI images of the fetus and placenta should be acquired in at least one plane. Notably, susceptibility sequences (EPI/T2*) may support the identification of bleeding and calcifications, and other sequences should be examined according to need.

As for the optimal timing, the main challenge is ongoing brain fetal development. Early gestation is characterized by incomplete structural maturation and poor anatomical differentiation, therefore an accurate prognosis is hard; additionally, enough time after the infection is needed for structural changes to occur [3].

According to the ECCI consensus, infected fetuses should undergo an MRI evaluation during the third trimester, with an additional postnatal MRI in symptomatic neonates.

Moreover, if the timing of vertical transmission is unclear or if the fetal infection derives from an MPI in the first trimester, further imaging is deemed appropriate [25]. In fact, other studies suggested using MRI at 32 weeks of gestation in addition to US examination, in order to improve the prognostic evaluation of cCMV; this combined strategy has shown a high negative predictive value for the presence of symptoms at birth and the development of moderate to severe outcomes [11].

Unexpectedly, Cannie et al. found that fetal MRI was equally effective in predicting SNHL and neurological damage at both 27 and 33 gestational weeks. Furthermore, Doneda et al. have demonstrated that CMV-associated brain abnormalities can sometimes be found before this gestational age, even at 25 week [3], [34]. In fact, anticipating disease detection is of outmost importance, also due to the abortion legal limit, which in many countries is 24 weeks of gestation [3].

Imaging findings

Given the neurotropism of the virus, prenatal imaging in cCMV is mainly focused on the brain. However, observable (macroscopic) damage is a late discovery, and it could evolve progressively or even regressively over time.

For these reasons, MRI results are often vague. According to Diogo et al., white matter signal abnormalities and ventriculomegaly are the most frequently reported CNS anomalies, although these findings are not specific to CMV infection. Some typical findings, although less commonly seen, are ventriculitis, intracranial calcification, as well as temporal lobe lesions (e.g. cysts, growth of temporal horns). Specifically, patients with cCMV frequently have ventricular abnormalities. Prognosis appears to be correlated with the degree of ventriculomegaly, with severe ventriculomegaly (>15 mm) typically linked to worse outcomes [3].

Due to the wide spectrum of imaging findings, MRI protocols should be tailored in order to carefully assess each abnormality. For example, T2W hyperintensities in the white matter may represent a pathological finding; however, in the third trimester they are often physiological, therefore their interpretation can be challenging; some studies reported a 17% false positive rate, higher than US, where all false positives regarded ambiguous white matter hyperintensities [3], [34], [35]. On the other hand, ventricular adhesions, also known as pseudocysts, can be recognized in T2W images as thin strands of tissue crossing the ventricle. T2W FLAIR sequences, which suppress the CSF signal and highlight nearby lesions, can help visualize them [3]. Figure 1 present an example with both findings detected through T2W sequences, at different gestational weeks.

Another CMV-specific finding is the presence of intracranial calcifications, which can also

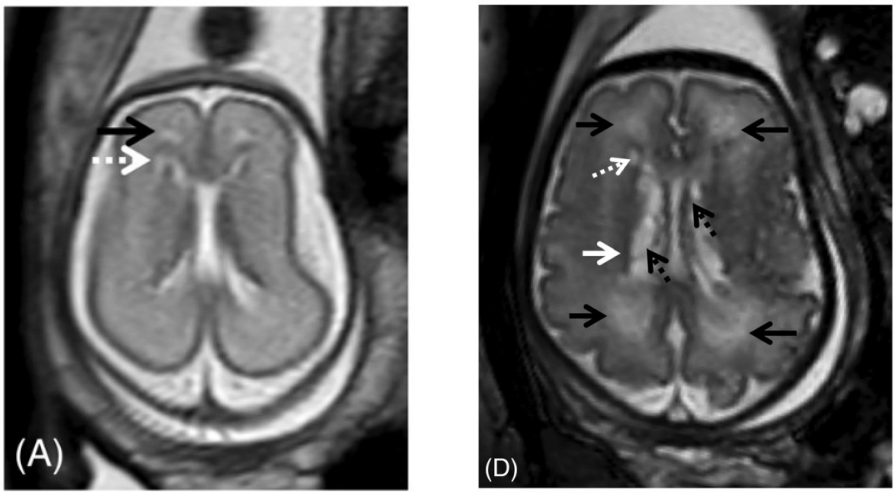


Figure 1. T2WI images of the same fetus, respectively at 24 and 34 gestational weeks. On the left, frontal periventricular caps (A, black arrow) and a small periventricular cyst (A, white dashed arrow). On the right, during follow-up, the cyst is still present (D, white dashed arrow), with other pseudocysts (D, black dashed arrow). Additionally, there are frontal and parieto-occipital white matter signal changes (D, black arrows) [3].

be seen in US. They can be identified through MRI as regions with low T2W and T1W signal, as well as low signal in echo planar imaging (EPI), but they are less noticeable if punctuated.

Generally, when the described lesions are identified in the anterior temporal lobe, the probability of a cCMV infection arises [3]. However, only a selection of the possible findings has been presented here; the complete list of brain findings is provided in Table 3.

Of note, cCMV can also involve other organs, including the placenta. Although there is incomplete literature on the matter, frequent findings are hepatomegaly and splenomegaly, as well as echogenic bowel and abnormal fluid collections [3]. These manifestations will be not discussed in this work, as the present study focuses specifically on CNS involvement.

Table 3. List of possible fetal MRI findings of the brain in cCMV infections (adapted from Diogo et al. [3]*).*

Finding	***MR characteristics***	***Notes***
WM hyperintensities	T2-hyperintense inhomogeneities of the WM Low SI on DWI Low SI on T1W/T2W FLAIR	Subjective Difficult to interpret particularly in the third trimester Temporal lobe correlates with worst prognosis
Ventriculomegaly	Increased lateral ventricle size (>10 mm), measured at the atria Mild: 10-12 mm Moderate: 12-15 mm Severe: >15 mm	May be uni- or bilateral Mild to moderate: low risk; severe: high risk of sequelae
Cysts/pseudocysts	Well defined lesions with SI similar to CSF on all sequences Most often periventricular	Inconsistent nomenclature temporal polar lesions highly predictive of CMV infection
Ventriculitis	T1w and T2w hyperintensity of the ventricular rim. On T2WI not visible due to juxtaposition to CSF; T2w-FLAIR useful if T1 is not informative	Rare finding. Most common lateral ventricles. Periventricular hyperechogenicity

Intraventricular septations/ adhesions	Tissue strands (T2w low SI) crossing the ventricles	Most common occipital horns
Cortical malformations/polymicrogyria	Cortical infoldings located in abnormal positions. Thickened cortical ribbon Blurry gray/WM margins on T2WI/FLAIR	MRI superior to US
Clefts (schizencephaly/ porencephaly)	Schizencephaly: transmantle cleft, lined by T2 hypointense (=cortex) ribbon81 Porencephaly: cleft with no cortical lining. Margins may show high T2w/FLAIR hyperintensity	Lesions secondary to disruption. Final manifestation depends on time of insult.
Calcifications	Low T2 and high T1 signal, often subtle Low T2*/EPI SI	Periventricular > deep gray nuclei > white matter
Cerebellar hypoplasia/dysplasia	Small vermis and/or hemispheres Increased infra/retrocerebellar space (megacisterna magna >8 mm) May have associated focal signal changes (i.e. hemorrhage, calcifications)	Rare fetal MRI Common postnatal imaging
Hippocampal dysplasia	Dilated temporal horns Verticalization of the hippocampal± internal temporal lobe atrophy	Often not described in fetal MRI. Common postnatal imaging. (DeVries)
Lenticullostriate vasculopathy	US diagnosis Low SI T2WI on basal ganglia Calcification (low EPI/T2* and high T1 SI) of basal ganglia	Late finding on MRI

However, the existence of brain abnormalities during pregnancy does not necessarily correlate with an unfavourable postnatal outcome, as previously clarified for WM hyperintensities in the third trimester. In this regard, no MRI severity scores based on pregnancy scans have been established to date with the aim of prediction postnatal impairment.

Furthermore, not even postnatal exams can completely rule out the possibility of late-onset disabilities; still, some progress has been made in this regard. In 2019, Lucignani et al. proposed an MRI severity score based on clinical signs detected in postnatal MRI scans, which proved to identify adverse outcomes (up to 2 years) with a sensitivity of 70% and specificity of 80% [36]. In 2025, Calandrelli et al. presented an innovative scoring method that included visual-semiquantitative MRI variables (e.g., volumetric assessments of brain regions) to grade the severity of brain damage; this integrated MRI score better reflected the degree of clinical impairment at follow-up [37].

In conclusion, MRI definitely plays a key role in the assessment of postnatal outcomes; however, there are significant limitations in terms of accuracy and timing of predictions.

1.2. Potential role of AI

In recent years, the application of AI in the medical field has grown considerably, ranging from prediction tools for auxiliary diagnosis to virtual assistants for targeted treatment [38], [39].

Across this wide spectrum, predictive modelling for early disease detection has quickly improved to support decision-making, because early diagnosis typically increases the likelihood of an effective therapy. This shift has been facilitated by the increasing number of electronic healthcare record (EHR) systems, since data from large databases, especially in standard format, is particularly well-suited for AI analysis. This also applies to imaging data,

where the implementation of the DICOM standard and Patient Archive and Communication Systems (PACS) have produced comparable advantages in imaging research [46].

While many predictive AI strategies for early detection have been explored for a variety of conditions, such as cancer, diabetes as well as congenital diseases [40], [41], there are still significant gaps in this area for cCMV, with few studies being conducted. A 2008 study by Boger et al. used an MLP (Multilayer Perceptron), trained on CMV genomic sequences obtained from amniotic fluids of infected women, to predict the disease's outcome. In 2024, Hederman et al. trained ML models on laboratory data (i.e., serological data) in order to find discriminative predictors for MPIs and NPIs [42], [43].

To our knowledge, no studies to date have investigated the topic of predictive modeling applied to fetal imaging in cCMV. Given the importance of an early diagnosis and prognosis in this context, their development could offer significant assistance in detecting subtle imaging patterns and enhancing early pregnancy risk stratification. This assumption is further supported by the fact that clinicians can occasionally recognize relevant imaging features on fetal MRI at this stage [34], thereby suggesting that meaningful diagnostic information may be already present.

The lack of studies on this topic may also be explained by the fact that, as far as we are aware, there are no publicly available datasets of clinical and/or imaging data for this condition, especially considering that follow-up should ideally be included to account for children with late-onset symptoms; however, efforts in this direction are ongoing [44]. On the other hand, due to the relatively low proportion of symptomatic cases, proprietary datasets are often limited in size. When combined, these restrictions underline the task's complexity, posing significant challenges for predictive modeling.

1.3. Objective of the thesis

Following these considerations, the objective of the present thesis is to conduct a feasibility study, aimed at evaluating the potential for detecting abnormalities in fetal MRI as early as 20 weeks of gestation. Furthermore, considering the explorative nature of the study, it is crucial to highlight its possible limitations, in order to ensure clarity and guide future research direction.

This investigation will be carried out using artificial intelligence, specifically Vision Transformer-base Masked Autoencoder (ViT-MAE) generative models [45], in order to evaluate whether MRI data acquired at 20 weeks contains sufficient information for accurate prediction of 30 weeks MRI scans' outcomes.

Two strategies will be explored: reconstruction-based anomaly detection using the ViT-MAE decoder reconstruction error, and classification based on latent representations. Their performance will be assessed to determine their robustness and potential for generalization in a larger-scale study. Therefore, the objectives also include verifying the stability of the training process, the quality of learned representations and the suitability of the extracted features for subsequent clinical tasks.

1.4. Related work

As previously noted, no research with this purpose has been published to date in the field of cCMV infection.

On the other hand, ViT-MAE models have already proven effective both within and outside of the medical imaging domain [46], [47], [48]. Proposed in 2021 by He et al., the ViT-MAE model is built upon a Vision Transformer architecture combined with a masking-based self-supervised strategy. Specifically, input images are split into patches of regular size; during training, a large portion of these patches is randomly masked, forcing the model to reconstruct them using only the visible ones. This masked-autoencoder framework can capture both local

and global structures and thereby learn meaningful and robust representations, without requiring large amounts of labeled data [45]. An extensive description of this architecture will be provided in Section 2.4.1.

In the medical field, explored tasks include multi-modal brain MRI representation learning with handling of missing inputs [46], as well as anomaly detection tasks using self-supervised MAE frameworks in breast MRI, such as the MAEMI model [49]. The latter aimed at localizing suspicious tumor regions with comparable performance to ordinary DCE-MRI subtraction images, which normally require the administration of a contrast injection. As an anomaly detection approach, the image reconstruction error of the image is expected to be higher in abnormal lesion regions, serving as an anomaly map that highlights potential lesions [49].

In 2025, Gupta et al. presented a MedMAE, a self-supervised ViT-MAE based backbone trained on LUMID, an extensive dataset with different imaging modalities (e.g. CT scans, MRI, X-rays) covering multiple anatomical regions. This model has been later finetuned on multiple medical imaging tasks, such as quality control, breast cancer and pneumonia prediction, as well as polyp segmentation in colonoscopies, with accuracy values ranging between 70.1% and 93.2% [48].

Overall, these findings show that ViT-MAE models have the ability to learn robust features from sparse annotated data, while also potentially providing spatial outputs. Such approaches have not yet been applied to fetal MRI for congenital CMV detection; however, by learning the underlying distribution of regular anatomical structures, the model may be able to identify deviations associated with potential abnormalities.

2. Materials and methods

This chapter outlines the study design, including the broader study framework, as well as the dataset description, preprocessing pipeline and model development.

2.1. Study design

The present thesis serves as a pilot study for a larger multicenter project that will explore the application of AI-models for the analysis of fetal brain MRI, with the goal of supporting early diagnosis and prognosis at approximately 20 weeks of gestation. The study follows a retrospective design, therefore all data were previously collected as part of routine clinical procedures and then anonymized for research purposes.

The overall project has several objectives. The primary aim is the evaluation of the diagnostic value of AI models, trained on 20-week fetal MRI scans, in predicting 30-week clinical outcomes, therefore enabling early diagnosis. In this setting, the 30-week outcome represents the study's primary endpoint, while the 20-week outcome offers an intermediate evaluation. Secondary aims will be early detection using postnatal MRI exams as the reference standard, localization and grading of fetal brain abnormalities on 20-week MRI scans, as well as investigating whether fetal MRI can act as an additional predictor together with standard clinical ones (e.g., gestational age at infection, viral load) for the prediction of sequelae in the postnatal period.

Therefore, the objective of this thesis is limited to the primary aim. This has been specifically adapted and narrowed to the application of promising ViT-MAE models on a limited data subset, before extension to the full multicenter cohort. In particular, the study investigates whether these models can be adapted to handle a reconstruction-based anomaly detection and a self-supervised classification task using latent representations.

This pilot investigation is motivated by the need to validate the proposed methodological pipeline prior to scaling to the full multicenter dataset. Therefore, crucial objectives include verifying the presence of significative predictive information in early gestation data, the

stability of the training procedure and the quality of learned representations for downstream clinical tasks.

Explorative analysis, model development, training and evaluation were all conducted on Python using the Visual Studio Code development environment; in particular, model development relied on PyTorch and related libraries.

2.2. Dataset description

2.2.1. Study population

The data for this project were retrospectively collected from two centres: Fondazione IRCCS Policlinico San Matteo (Pavia) and Vittore Buzzi Children's Hospital (Milan). The full dataset consists of fetal brain MRI examinations acquired at 20 and 30 weeks of gestation and, when available, postnatal scans, along with their corresponding neuroradiological reports and laboratory data.

Inclusion criteria for the study population were maternal age equal to or above 18 years, confirmed CMV infection via amniocentesis irrespective of maternal treatment, early timing of infection (i.e., during the first trimester) and presence of both 20- and 30-week scans for each patient. Twin pregnancies, preterm birth or congenital abnormalities not attributable to cCMV infection resulted in exclusion from the study.

As previously mentioned, the proposed approach focuses on a subset from the second centre, to evaluate its feasibility; the study population therefore consists of a total of 76 cases.

2.2.2. Data exploration

For each subject, two MRI examinations are present, acquired respectively at 20 ± 2 and 30 ± 2 weeks of gestation. Each scan consists of multiple sequences (e.g., SSH-TSE, bSSP, DWI), including the survey sequences; however, not all of them are available for each subject. Generally, there are more than one acquisition for each series. As explained by clinicians,

repeated acquisitions are usually required due to suboptimal fetal positioning, or poor image quality derived from fetal motion.

Examinations were performed using several different scanners: Ingenia, Achieva, Achieva dStream and Intera Achieva by Philips Medical Systems, as well as MAGNETOM Sola and MAGNETOM Vida by Siemens Helthineers. Due to this variability, the number of slices varies across subjects, and not only sequences and planes; additionally, out of 152 examinations, only 9 were performed at 3T, while the remaining 143 were acquired at 1.5T.

Every scan is associated to a neuroradiological written report. This can either describe the examination as normal (e.g., "normal", "within normal limits", "unremarkable"), provide a detailed statement reporting all parameters, not necessarily within the normal range (e.g. biometric parameters below the normal range), or report clear potential pathological findings (e.g., "hyperintensities in the T2-w series, presence of a pseudocysts"). In the absence of postnatal outcomes, which would represent the definitive reference, 30-weeks MRI assessments will serve as the reference standard, given that the pathology typically becomes more evident over time.

2.3. Preprocessing

The preprocessing pipeline included multiple steps: (i) sequences and planes' selection, (ii) slice trimming, (iii) data standardization, (iv) outcome coding and (v) oversampling.

2.3.1. MRI data selection: sequences and planes

In order to avoid model bias caused by the small and unrepresentative 3T scans' subgroup, these were not included in the final subset, due to significant imbalance compared to the amount of 1.5T scans.

Furthermore, since not all sequences were available across subjects, a sequence selection was needed. Naturally, the survey sequences, also known as *localizers* or *scout scans*, were

excluded from the analysis, since they are low-resolution, non-diagnostic acquisitions primarily used for scan planning. Among the others, the most prevalent sequence across all instances was the T2-weighted SSH-TSE, which was therefore selected for this study. Because of its high contrast resolution and robustness to motion, which justifies its constant presence as the standard imaging technique in fetal MRI, it is probably the best option for preserving useful information in a limited dataset.

Following this step, a careful visual inspection was conducted to select the highest-quality SSH-TSE series for each patient, due to the presence of repeated acquisitions. In the absence of a skilled clinician available at all times, when ambiguity arose, the most recently acquired series was selected, under the assumption that it would reflect a better acquisition than previous repeated attempts. However, some subjects still presented low-quality images, which led to their exclusion.

In order to minimize inter-subject variability and maintain a consistent focus on the fetal brain, only the axial plane was considered in this study. Compared to other orientations, as shown in Figure 2 and Figure 3, the axial view offers more standardized anatomical coverage and more consistent background intensity [50], while the sagittal view can be more heterogeneous because of variable fetal positioning and inclusion of non-brain moving structures (such as limbs); specifically, the exact midsagittal slice was indicated also by Zhao

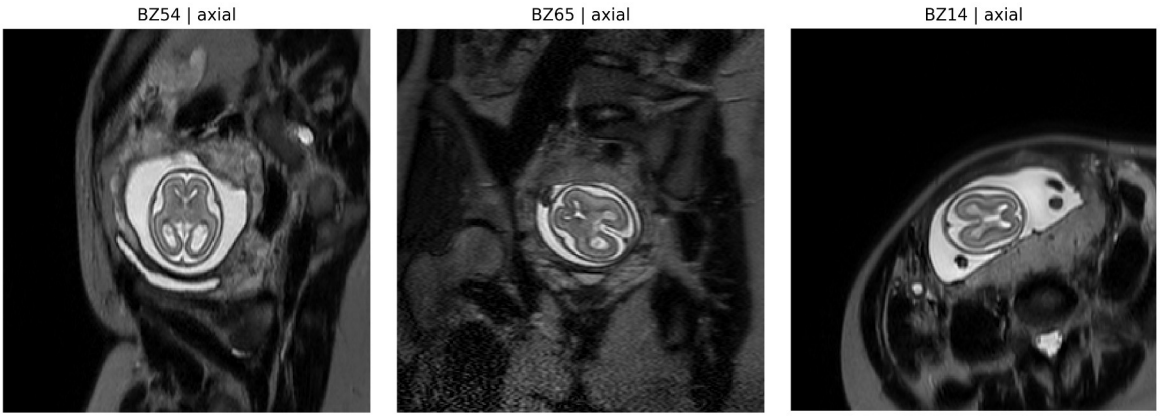


Figure 2. Central slices of three patients in the axial plane; the fetal brain is clearly visible, with a partial inclusion of the placenta.

et al. as often challenging due to the aforementioned reasons [51]. Restricting the study to the axial plane maximizes the already small percentage of relevant brain content; this is further supported by the lack of a brain segmentation step, which would otherwise enable total focus on the brain.

A simple alternative to brain segmentation would be image cropping; however, this approach would require an impractical manual definition of the ROI (region of interest), because of high variability in the spatial location of the fetal brain across slices and subjects. For this reason, in this initial pilot study this solution was not taken into account, also in order to assess whether limiting the analysis to the axial plane alone is sufficient.

As a result of the indicated procedures, the number of subjects was reduced to 66.

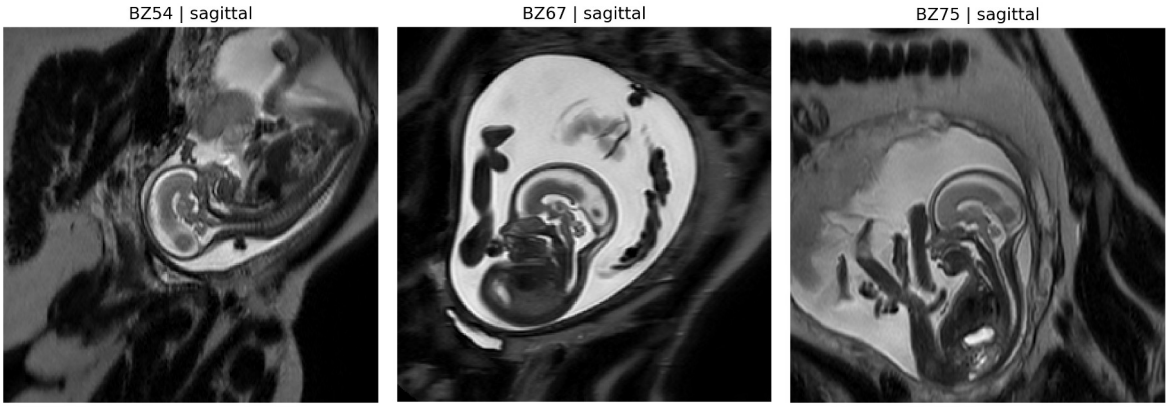


Figure 3. Central slices of three patients in the sagittal plane; while the whole brain is clearly visible, the background is much more heterogeneous., From left to right: fetus with clearly visible spine and not spinal cord; fetus with part of the stem and part of the legs visible; fetus with all limbs clearly identifiable and whole spinal cord.

2.3.2. Slice trimming and data standardization

Another issue that surfaced during the visual inspection was the imbalance in the number of slices, despite the identical sequence. In fact, the adapted ViT-MAE model requires an equal number of slices for each item, therefore unnecessary slices had to be eliminated.

In order to achieve this, a simple assumption was made: central slices are considered the most informative, since border slices usually involve external and less important structures. The strategy adopted therefore included keeping central slices and trimming the border ones, until the minimum slice number (7) was reached.

Data standardization was carried out after this point. Before its application, the average pixel intensity histogram (Figure 4) revealed a distribution ranging from 0 to 1085, with a prominent peak at 0, clearly attributable to background pixels.

We therefore compared two methods: a simple standardization based on mean and standard deviation of intensity values, and a foreground-based standardization, which involved the application of a mask to exclude background pixels (which were set to -3.0) and computing standardization on the foreground values.

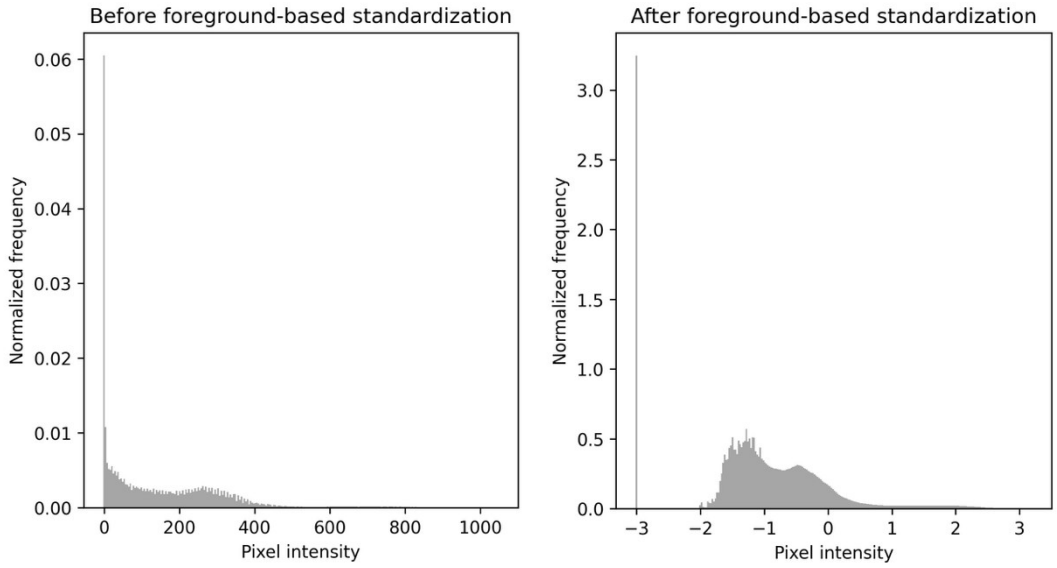


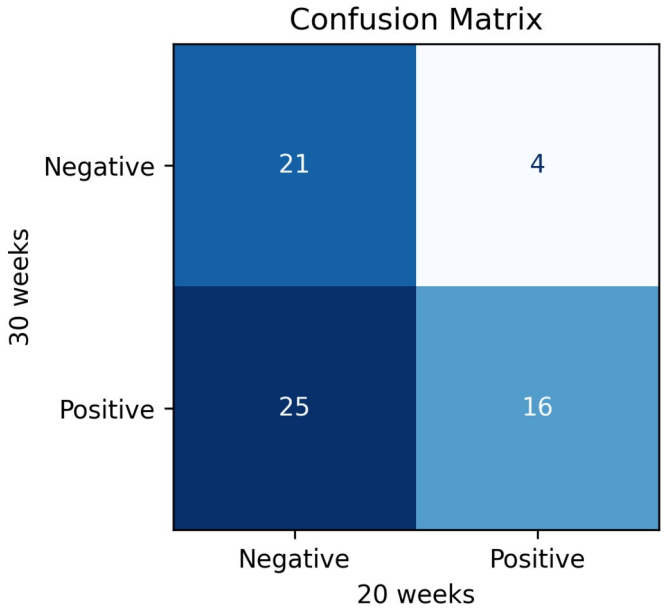
Figure 4. Average pixel intensity histogram in the axial plane, across slices and subjects in the 20-week MRI scans, respectively before and after foreground-based standardization. In the latter, background pixels are set to -3.0.

However, due to the variability of intensity ranges across subjects, the standardization was applied separately on each patient.

2.3.3. Outcome definition

To cast the problem as a classification task, neuroradiological reports were converted into a dichotomized coded value for the outcome. In the absence of a qualified clinician for detailed annotation, we adopted a simple binary coding strategy: reports explicitly stating "Normal/Within the normal limits" were assigned a negative label, whereas any other deviation was given a positive label. Clearly, this represents a strong assumption, because some findings classified as "abnormal" may actually be physiological variations (e.g., biometric measurements within the lower or upper range of normality, hyperintensities in the third trimester).

This categorization was applied to both the 20-week outcome and the 30-week outcome, obtaining the confusion matrix shown in Figure 5. As previously mentioned, the 30-week outcome serves as reference, whereas the 20-week outcome is an intermediate predictive assessment.



Metrics	**Percentage**
Sensitivity	39%
Specificity	84%
Positive predictive value	80%
Negative predictive value	46%
Overall accuracy	56%

Figure 5. On the left, confusion matrix between 20-weeks outcomes (grouping variable) and 30-week outcomes (ground truth); on the right, the corresponding performance metrics, including sensitivity, specificity, positive and negative predictive value, and overall accuracy.

Of the 66 individuals, 37 had a stable diagnosis (either negative or positive), whereas 25 previously thought negative were later classified as positive. On the other hand, only 4 previously considered positive were later diagnosed as negative.

These findings indicate low sensitivity (39%) and high specificity (84%), demonstrating that clinicians are good at identifying true negatives, whereas they struggle to detect true positive cases. This is particularly evident in the high number of false negatives (25), suggesting that many patients who were initially classified as negative later evolved into detectable positive cases by the 30th week. However, this is likely not entirely attributable to human error, because the condition may not have been sufficiently developed or visible at the earlier time point.

Clinicians rarely overdiagnose the illness, since the positive predictive value is rather high (80%); this means that when abnormalities are identified, they are usually reliable. Conversely, the low negative predictive value (46%) highlights the uncertainty associated with early negative assessments. Overall, these findings imply that diagnostic accuracy is strongly impacted by disease progression over time.

As already stated, the model was trained on two tasks: reconstruction-based anomaly detection; and classification task. The dataset was split into training, validation and test sets using an 80/10/10 and 70/15/15 ratio, respectively.

In both settings, the 30-week outcome served as the ground truth, since it represents the definitive clinical outcome in the absence of a postnatal follow-up. The 20-week outcome was instead regarded as grouping variable for the inputs, since the goal is to determine whether information at 20 weeks is predictive of the later outcome.

For the anomaly detection task, only scans from the 20-week negative class (majority class) were used for training, for a total of 28 samples in the training set, in order to learn a representation of normality and later identify those who deviate from the norm. The validation

set consists of 9 negative subjects, while the test set includes the remaining 9 negatives and all 20 positives, for a total of 29 samples.

For the classification task, both classes were included during training. However, given the strong imbalance between the two groups in the training set (30 negatives vs 16 positives), oversampling was applied to the training set to equalize class distribution.

In particular, two oversampling strategies were evaluated: simple oversampling, which involved replications of random MRI scans from the minority class, and data augmentation-based oversampling, in which rotated MRI scans of the minority class were added to the training set.

Finally, all images were resized to 224 x 224 pixels, which is the required fixed size for the input layer of the model architecture, which will be described in the following section.

2.4. AI models

2.4.1. Original Vit-MAE architecture

The proposed architecture is an adaptation from the ViT-MAE model (base version) presented by He et al. in 2021 [45]. The original framework involves an asymmetric encoder-decoder architecture, as reported in Figure 6: the encoder receives only a visible subset of patches from a partially masked image, while the decoder reconstructs the original image from the latent representation and mask tokens. This enables the model to speed up training and increase accuracy by forcing it to learn from few patches and generalize well.

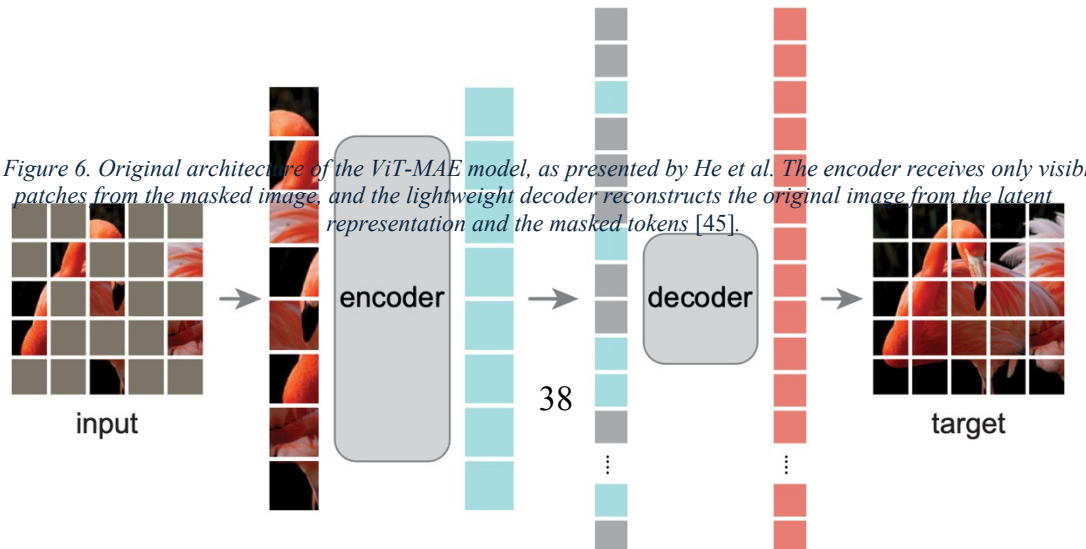


Figure 6. Original architecture of the ViT-MAE model, as presented by He et al. The encoder receives only visible patches from the masked image, and the lightweight decoder reconstructs the original image from the latent representation and the masked tokens [45].

Several pretrained versions characterized by different architectural scales (e.g., ViT-Base, ViT-Large, ViT-Huge) are available and can be accessed through platforms such as the HuggingFace model hub. Our adaptation derives from the ViT-Base configuration, which is characterized by the lowest number of parameters, as well as reduced depth, embedding dimension (width) and number of self-attention heads.

Typically, the model operates on input images of size 224 × 224, which are then split into 196 non-overlapping 14 × 14 patches. A key component of the masked autoencoder (MAE) is the application of a high masking ratio, commonly 75%: this means that a large portion of the patches is randomly removed, before being processed by the encoder. It has been demonstrated that a high masking ratio can optimize accuracy while reducing training-time and memory consumption.

Subsequently, each visible patch is linearly projected into a fixed-dimensional patch embedding vector, combined with positional encodings to enable correct spatial positioning during reconstruction. Therefore, only the resulting sequence of visible patch tokens is passed to the encoder, whereas mask tokens are skipped.

On the other hand, the lightweight decoder receives the encoded visible tokens along with a set of newly generated mask tokens, that replace the missing patches; positional embeddings for all patches (visible and masked) are included as well, in order to preserve spatial information. Through a series of transformer layers, the decoder returns the reconstructed full image.

The model is trained by minimizing the mean squared error (MSE) between the reconstructed image and the original input, computed only over the masked patches. This implies that learning focuses only on missing content, rather than copying the already visible regions. As a consequence, the model learns to infer global structure; however, the trade-off is a poor reconstruction of the visible patches, as illustrated in Figure 7.

This reconstruction task can be considered self-supervised learning, since the model learns without requiring explicit labels. As a result, the MAE architecture is able to extract high level latent representations, which can be used for transfer learning.

Specifically, after the pre-training phase, the encoder can be reused as a feature extractor. In this stage, the learned decoder weights can be frozen or fine-tuned, and connected to a classification head, which uses the encoder embeddings as inputs. In their work, He et al. evaluated transfer learning across multiple downstream applications, including object detection and segmentation, as well as image classification; their results demonstrated that ViT-MAE can outperform traditional supervised pre-training models.

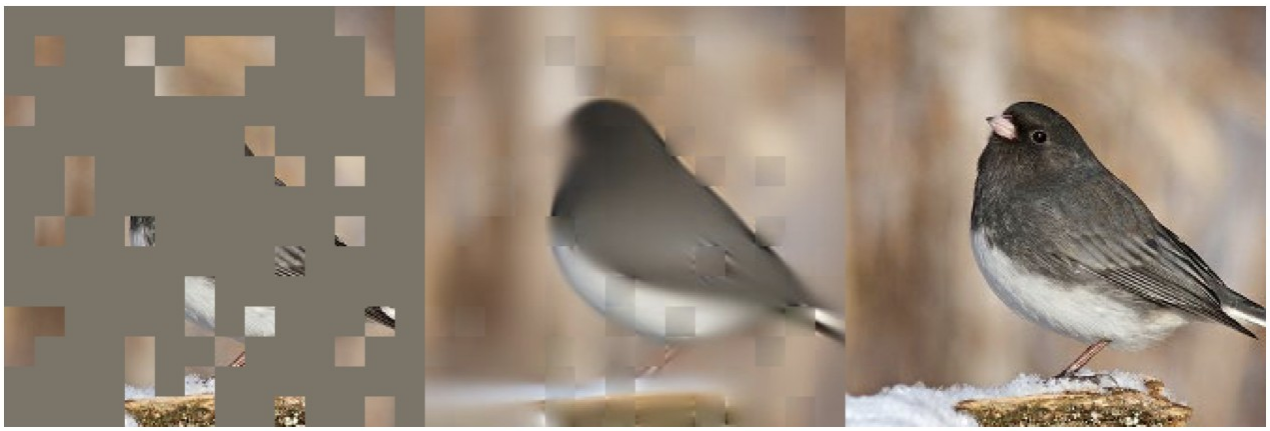


Figure 7. Example results of an image from ImageNet: masked image with an 80% mask ratio (left), MAE reconstruction (middle), and ground-truth (right). It is evident that visible patches are reconstructed less accurately compared to hidden ones. Figure reproduced from He et al. [45].

2.4.2. Proposed framework

In this study, two complementary strategies have been investigated.

The first approach is based on the assumption that deviations from learned normal anatomical patterns can indicate abnormal cases. Consequently, an anomaly detection framework was developed, in which the model is trained on “normal” data. Specifically, image reconstruction is performed using the full ViT-MAE architecture, including its decoder

component, with appropriate modifications. Therefore, regions or samples that exhibit higher reconstruction errors should reflect potential anomalies.

Second, a classification task is performed using latent representations extracted from the pretrained ViT-MAE encoder, which are anticipated to encode high-level anatomical and contextual information of the fetal brain. These representations serve as input to a classification head, which will be a Multi-Layer Perceptron (MLP).

To adapt the original model to these strategies, several modifications were needed. First, the input layer normally receives one image at a time through its RBG channels. In our adaptation, given that each patient's scan consists of multiple grey-scale slices, color channels were transformed into slices' inputs. Specifically, the number of channels was increased to 7, reflecting the number of slices for each patient; therefore, each patch embedding is a representation of the entire image series for a specific spatial region.

Masking ratio remained unchanged (0.75). Then, other modifications were applied differently depending on the approach; they will be therefore described separately in the next paragraphs.

Anomaly detection

In this strategy, training is performed only on the 20-week negative cases (majority class), so that the model can learn to generate strong latent representations of the normal class). Patch size was first set to 16 × 16 and then reduced to 8 × 8 in order to capture finer details, which was necessary given the limited size of fetal brain region compared to the whole image.

Unlike the original model, loss is computed measuring the reconstruction error including both hidden and visible patches, forcing the model to learn from both to get better reconstructions. The MSE remains the metric in use.

After pre-training, validation loss is used to evaluate learning performance. Additionally, the test set is reconstructed to assess whether there is a significant difference in model reconstruction between the 30-week positive and the negative subgroups. Our goal is, in fact, to find out whether there is enough imaging information at 20 weeks that could already predict the later outcome. If this assumption is correct, the model is supposed to exhibit worse reconstruction on positive scans present in the test set.

Classification task

In the alternative approach, pretraining is performed on both 20-week classes, allowing the model to learn strong generalizable representations. As previously stated, because of oversampling, the training set includes an equal number of positive and negative cases (30 each).

Patches' size was directly set to 8 × 8. The reconstruction error is computed as in the original model, therefore only on hidden patches; then, MSE is minimized.

After pre-training, the encoder weights are frozen, and a self-supervised finetuning operation is performed by attaching a classification head to the encoder. This head consists of a simple MLP, that receives the latent representations and predicts 30-week outcomes.

An explorative evaluation was conducted prior to finetuning, by projecting latent representations into a reduced dimensional space through PCA and UMAP. The goal was to identify potential clustering between the two 30-week classes, which could indicate the presence of diagnostically discriminative information. The latent representations for each patient can be obtained by averaging patch embeddings across each scan, or CLS tokens, special learnable vectors designed to aggregate global information across all patches thanks to the self-attention heads. Therefore, the analysis was performed separately on both representations.

Additionally, the analysis was applied to both the training and test set, since the latter was too small to reliably detect any significative cluster.

Subsequently, during fine-tuning, the model was optimized using the cross-entropy loss function on the training set, which measures the discrepancy between predicted probabilities and ground-truth labels and strongly punishes confident mistakes. The performance was then monitored on the validation set, to guide hyperparameter tuning and select the best-performing model based on validation loss.

3. Results

The following sections report the results obtained from the proposed methodology, separately for each of the two developed approaches. Several combinations of the different preprocessing strategies and model parameters were evaluated; however, only those of note have been reported in the present document.

3.1. Anomaly detection

3.1.1. Best model selection: training and validation losses

As previously stated, models were trained for the reconstruction task with image patches of varying sizes, specifically 16 × 16 and 8 × 8. As a result, Table 4 presents the lowest validation loss obtained under these conditions, combining also various standardization strategies and model hyperparameters, such as learning rate (LR) and minimum learning rate; the related curves for training and validation loss are shown in Figure 8 and Figure 9.

In the table and figures, model names follow the structure: **[PatchSize]p_[Standardization type]_[Configuration];** W stands for whole-image standardization, whereas F indicates foreground-based standardization. Same configuration number indicates same LR and minimum LR parameters.

For all models, the batch size was set to 10 samples, and training was performed for a maximum number of 600 epochs, with early stopping applied if the validation loss did not improve for 10 consecutive epochs (patience = 10) or if the predefined maximum time training time was reached. In addition, a learning rate scheduling strategy was employed: when the validation loss plateaued, the learning rate was reduced by a 0.1 factor, until the minimum LR was reached.

Table 4. List of most relevant models along with hyperparameters, preprocessing steps and minimum validation loss. Patience and batch set are set to 10 and 8, respectively. Model name follows the structure [PatchSize]_[Standardization type]_[Configuration]; W stands for whole-image standardization, whereas F indicates foreground-based standardization. Same configuration indicates same hyperparameters.

Model	Patch size	Standardization	LR	Min LR	Min. validation loss
16p_W_0	*16 × 16*	*Whole image*	*0.005*	*1e-08*	***0.2949***
16p_W_1	*16 × 16*	*Whole image*	*0.001*	*1e-08*	*0.2997*
16p_F_0	*16 × 16*	*Foreground*	*0.005*	*1e-08*	*0.9575*
8p_W_1	*8 × 8*	*Whole image*	*0.001*	*1e-08*	***0.1958***
8p_W_0	*8 × 8*	*Whole image*	*0.005*	*1e-08*	*0.7156*
8p_W_2	*8 × 8*	*Whole image*	*0.001*	*1e-09*	*0.2912*
8p_F_2	*8 × 8*	*Foreground*	*0.001*	*1e-09*	*0.5067*

Since there are only nine patients in the negative subset (due to the limited size of the negative subset described in section 2.3.3) the presented metrics may not be entirely reliable and should be interpreted with caution.

Overall, except for the foreground-normalized models, the validation loss curves exhibit an initial exponentially decreasing trend, followed by a plateau phase with minimal oscillations. Therefore, almost all models early converge to a suboptimal minimum, mostly in a range between approximately 0.3 and 0.2: despite the initial decrease in loss, further learning appears to stall.

Configurations with whole-image standardization generally achieve significantly lower minimum validation loss, consistently outperforming foreground-based standardization across comparable conditions. In particular, foreground normalization results in markedly higher validation loss values (0.9575 and 0.5067 for the minimum losses).

Additionally, the results also indicate that increasing the learning rate to 0.005 does not always improve performance, and may even worsen it (e.g., 0.7156 for Model 8p_W_0).

Among the reported configurations, the best performance is achieved with a patch size of 8 × 8, a learning rate of 0.001 and whole-image standardization; this configuration reaches a minimum validation loss of 0.1958, which outperforms the other models. This suggests that

the combination of smaller patch sizes and whole-image normalization better preserve relevant structural information for the task.

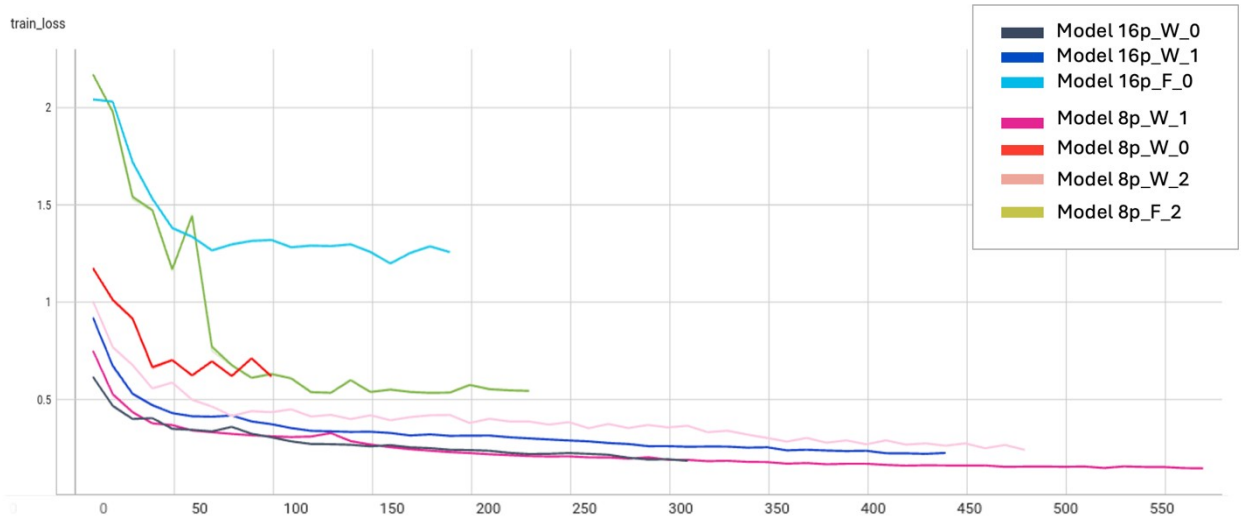


Figure 8. Comparison of training losses' curves of the most relevant model. The y-axis represents the training loss, whereas the x -axis reports the number of training steps.

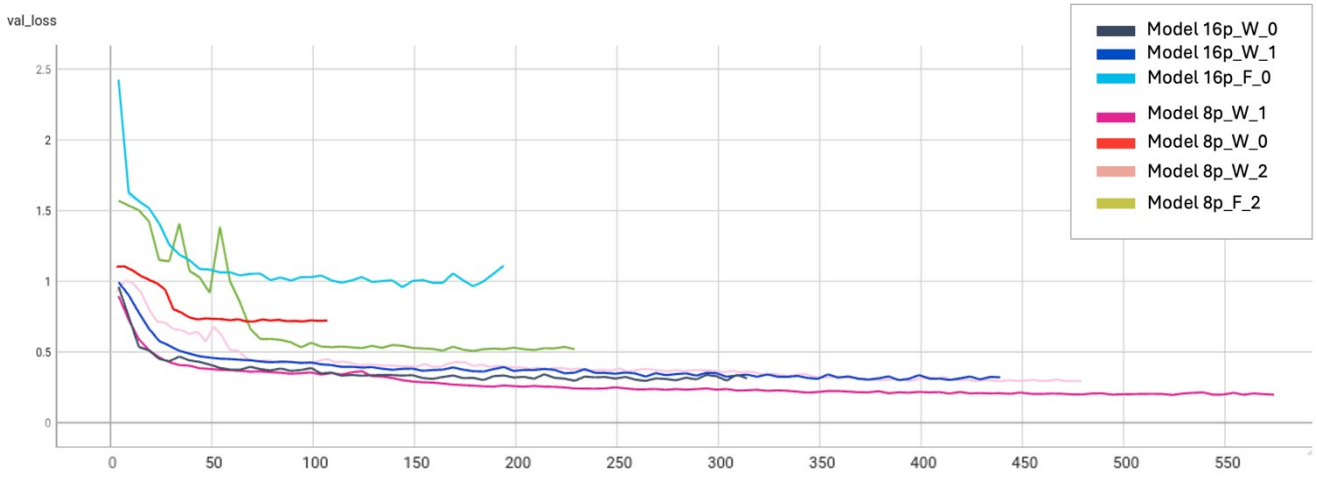


Figure 9. Comparison of validation losses' curves of the most relevant models. The y-axis represents the validation loss, while the x-axis reports the number of training steps.

3.1.2. Comparison of reconstructions with different patch sizes

Smaller patch sizes are reportedly better for capturing finer details. Therefore, an additional anecdotical comparison between the top-performing 16 × 16 model (Model 16p_W_0), and the best-performing 8 × 8 was conducted on a randomly selected positive sample: the results for the central slice are presented in Figure 10, along with the original image.

From the presented results, the model with 16 × 16 patch size seems to be capable of capturing well the global structures, including general boundaries and intensity distributions;

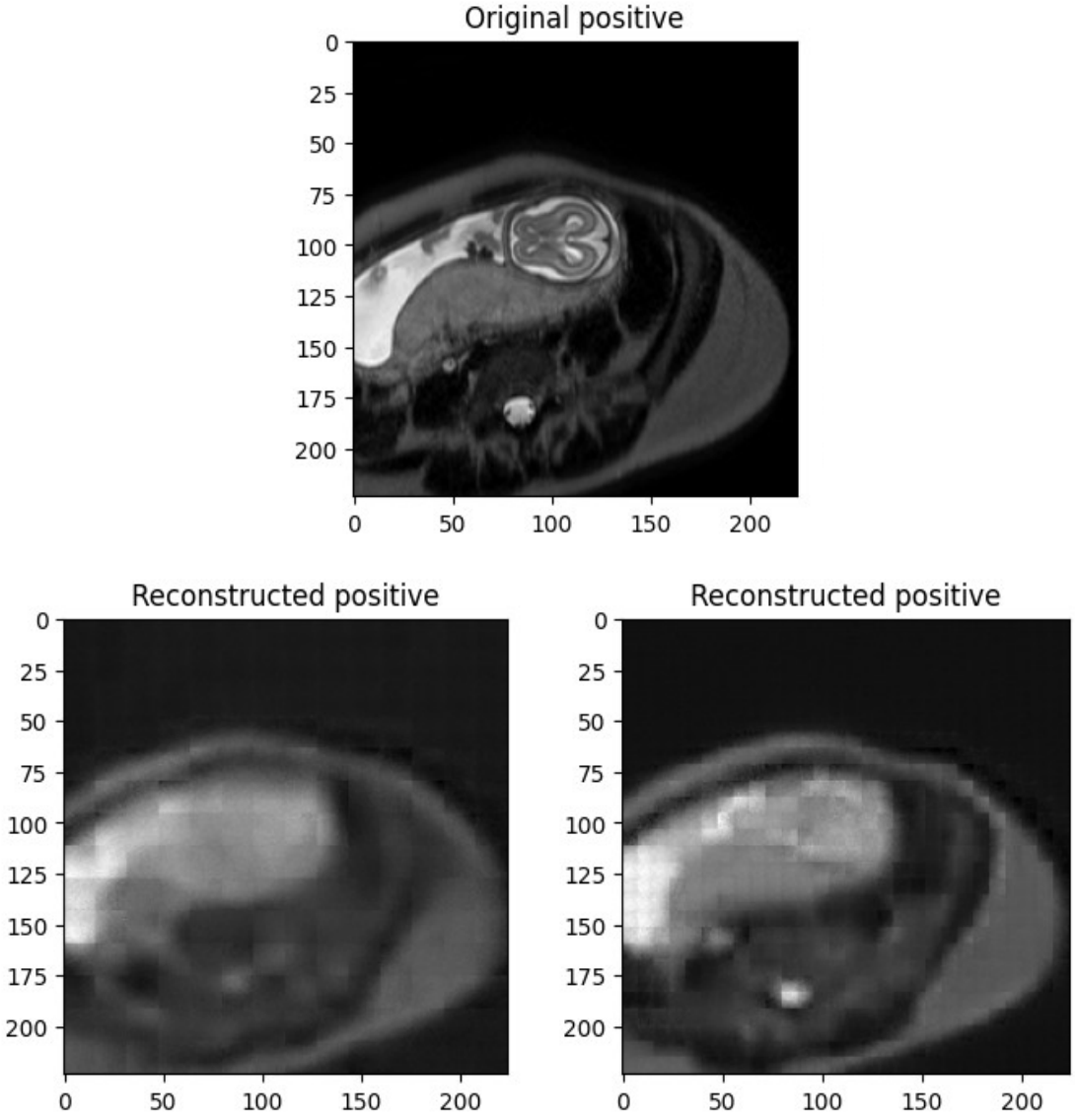


Figure 10. Comparison of models based on different patch sizes applied to the same patient. Above, the original central slice; below, the reconstructed images with 16 × 16 (left) and 8 × 8 (right) patches.

however, it lacks sensitivity for finer anatomical details. In contrast, the model with smaller-size patches performs noticeably better, showing improved delineation of subtle structures both inside and outside the placenta. For instance, the 16 × 16 model fails to detect the spine and struggles to identify the grey, arch-shaped structure close to the fetal brain.

3.1.3. Reconstruction error visualization

The best-performing model was subsequently applied to the entire test set.

For each of the two classes, the samples with the lowest and highest reconstruction error, measured in terms of MSE, were selected for further analysis. Since reconstructions are performed on the entire MRI volume for each patient, the comparison focused on the central slice of both the original and reconstructed MRI scan.

In particular, for each selected case, the original slice is compared with the related reconstruction, along with a heatmap of the slice reconstruction error, in order to highlight the regions with the greatest discrepancy. The resulting visualizations are presented in Figure 11, Figure 12, Figure 14 and Figure 15.

Furthermore, the **average** reconstruction error was computed across all slices and displayed once more as heatmaps to provide a more comprehensive analysis. The best-case scenario is depicted in Figure 13, while the worst-case scenario is shown in Figure 16.

Across the presented reconstructions, errors are generally minimal in background regions, which are predominantly black. This is expected, because low-intensity areas are typically more uniform and therefore easier for the model to reproduce accurately. This observation may explain why the MSE for the positive sample in the best reconstruction (0.1240) is lower than that of the negative sample (0.1906); as it can be seen in Figure 13, the negative sample (left) presents a relatively small black area, compared to the wide black region of the positive sample (right). A similar trend is also observed in the worst-case reconstructions, where the subgroups' performance is reversed: the positive sample has a poorer MSE (0.2298) with few

restricted background area, whereas the negative sample performs marginally better (0.2059), likely due to the presence of slightly more frequent black regions, mainly within maternal body.

In contrast, maternal organs exhibit moderate and more spatially distributed reconstruction errors. However, their edges are usually well-reconstructed, suggesting that the model captures structural features and edges more effectively than subtle intensity variations inside this area.

Furthermore, there is a noticeable concentration of other moderate errors in the fetal brain, but the highest ones are observed within or along the placenta. A likely explanation lies in the higher signal intensity in these regions: since they are closer to white (i.e., higher intensity values), even slight absolute deviations can lead to bigger errors. Additionally, these regions are usually characterized by more complex textures and variability, which makes them inherently more challenging to reconstruct accurately. In fact, high reconstruction errors can also be found in black small spots, e.g. in the fetal brain in Figure 11, likely due to their positioning in an otherwise bright area such as the placenta.

Best reconstructions among classes

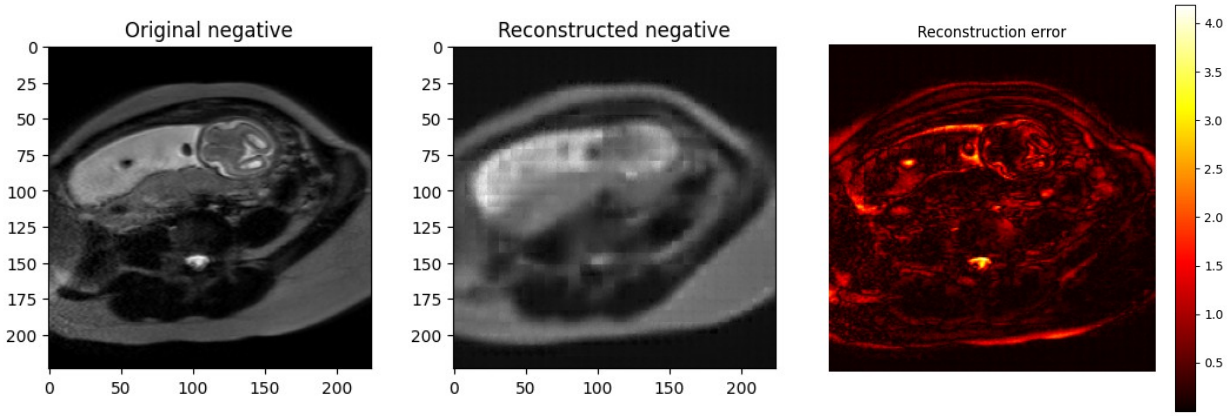


Figure 11. ***Best*** *reconstruction for the* ***negative*** *subgroup: original image (left), reconstructed image (middle) and reconstruction error (right). Errors are mainly found along boundaries; two small areas corresponding to a black spot and a vertebra present the highest errors.*

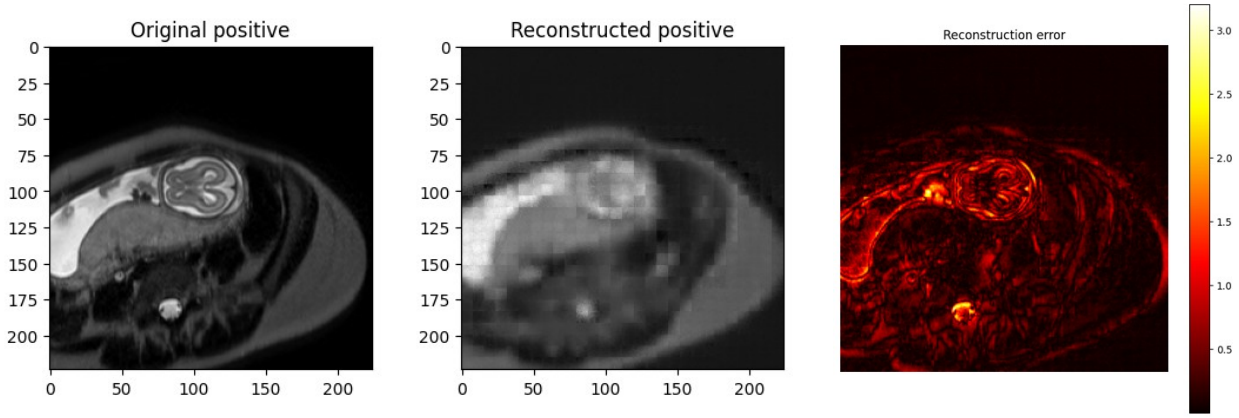


Figure 12. ***Best*** *reconstruction for the* ***positive*** *subgroup: original image (left), reconstructed image (middle) and reconstruction error (right). The highest errors (yellow-white areas) are found along the edges of placenta and fetal brain.*

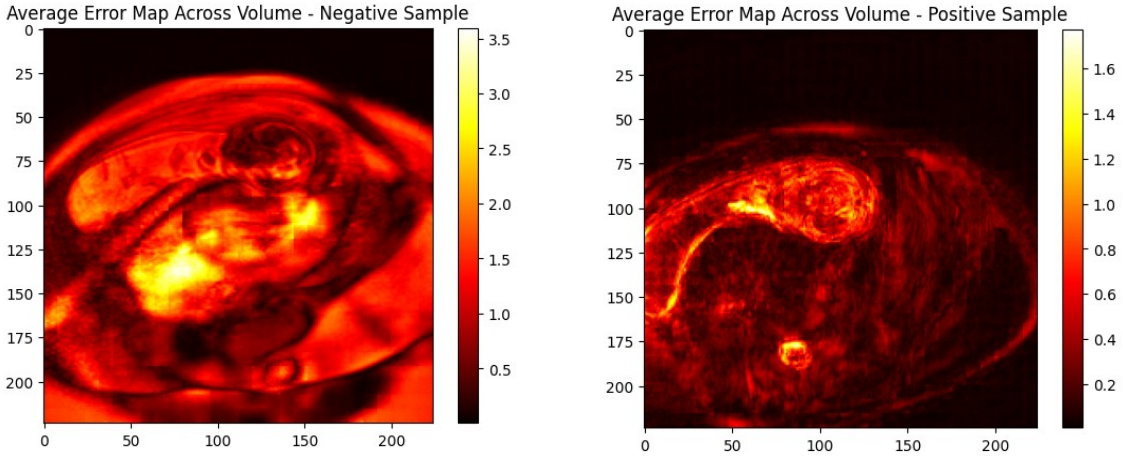


Figure 13. Heatmaps of average reconstruction error across the slices of each patient, respectively for the negative (left) and positive (right) best reconstructed sample.

Worst reconstructions among classes

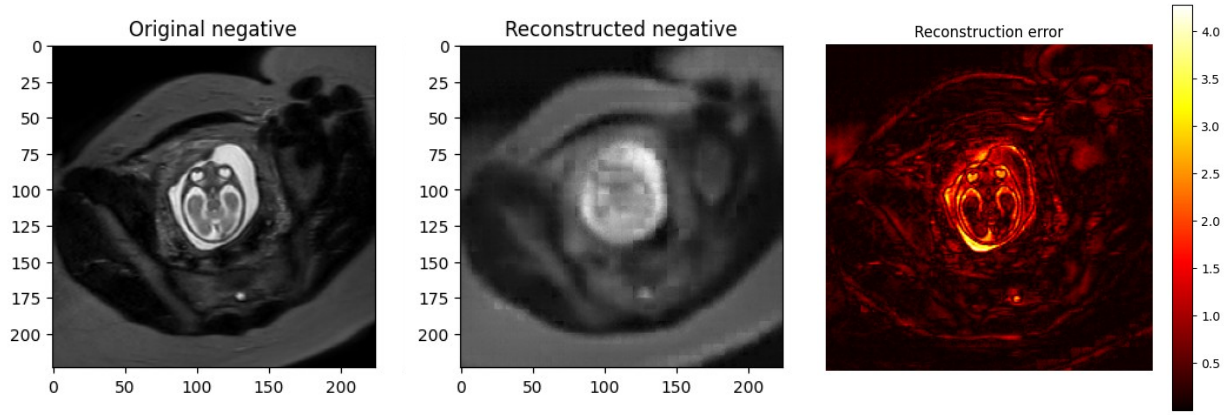


Figure 14. ***Worst*** *reconstruction for the* ***negative*** *subgroup: original image (left), reconstructed image (middle) and reconstruction error (right). In the heatmap, yellow-white areas present the highest errors, showing one inside the fetal brain.*

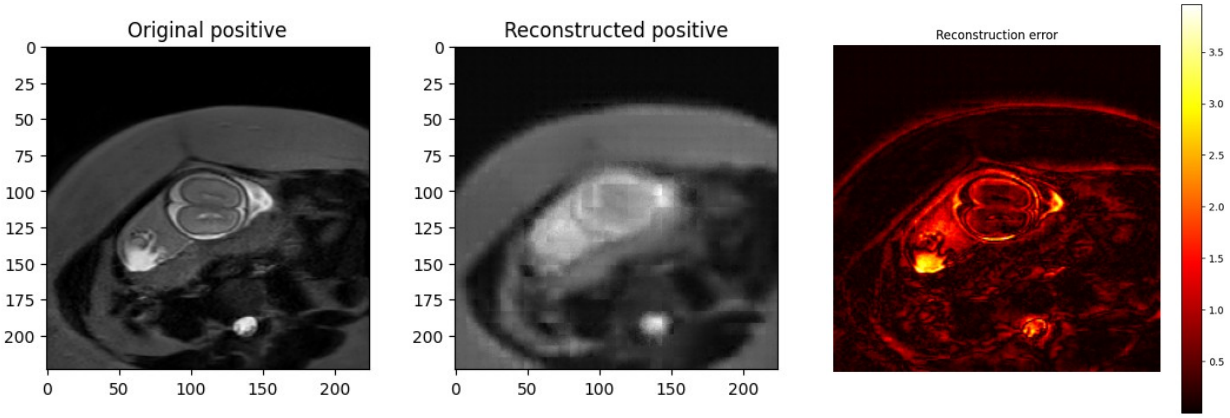


*Figure 15. **Worst** reconstruction for the **positive** subgroup: original image (left), reconstructed image (middle) and reconstruction error (right). In the heatmap, highest errors are found in the amniotic fluid.*

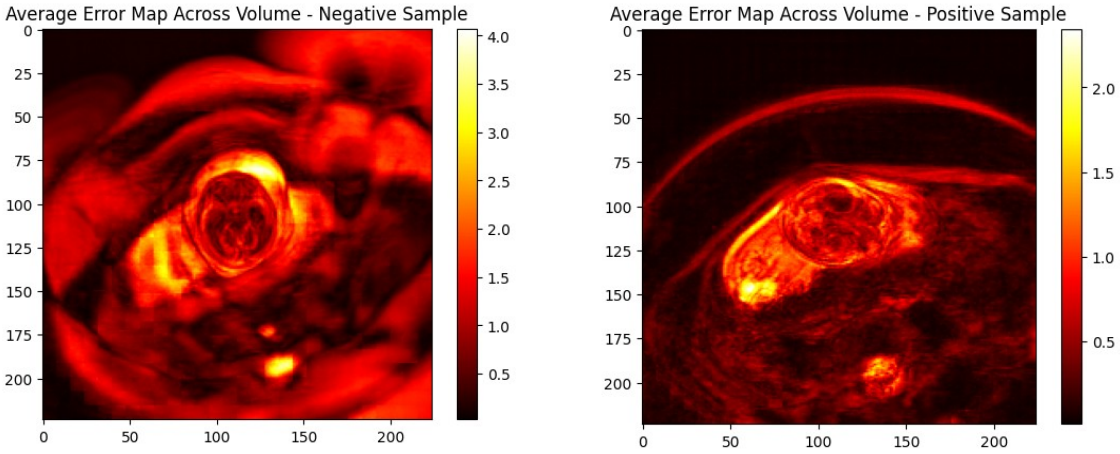


Figure 16. Heatmaps of average reconstruction error across the volume of each patient, respectively for the negative (left) and positive (right) worst reconstructed sample.

3.1.4. Quantitative assessment of reconstruction performance

Figure 17 displays the distribution of reconstruction errors across subjects, stratified by error magnitude; this was computed separately for each class.

While the negative class appears to be concentrated in a small number of higher-error regions, the positive class exhibits a more dispersed distribution across error bins. However, as previously discussed, this pattern may be influenced by an unfavorable distribution of the few negative samples, particularly in relation to the presence of small background regions which lead to worse performance.

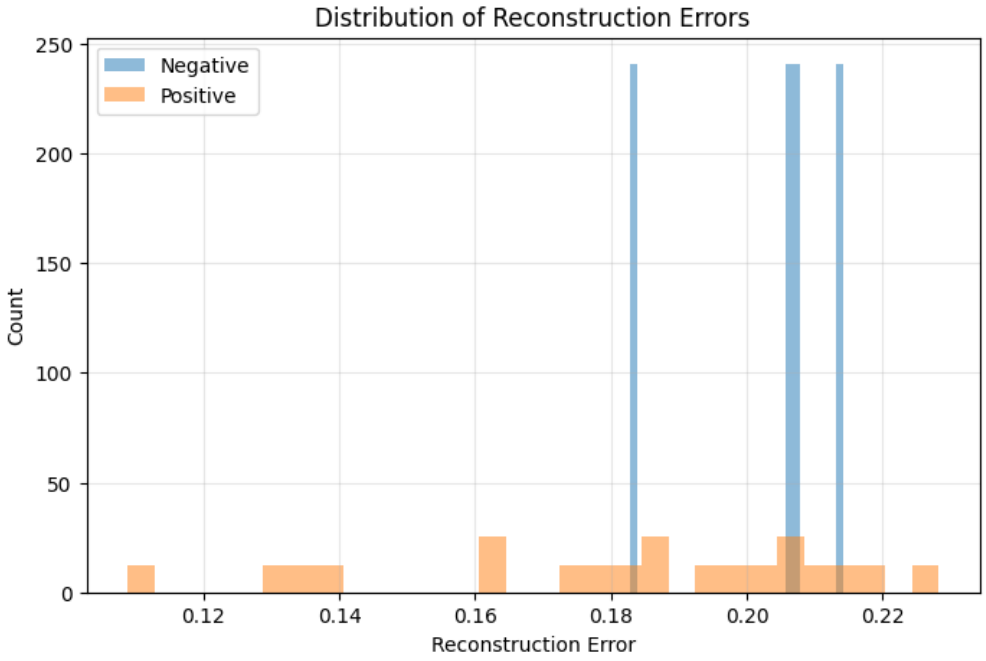


Figure 17. Distribution of reconstruction errors.

3.2. Classification task

3.2.1. Selection of the best pretrained model

Several models were pretrained, in order to get meaningful embeddings for the downstream fine-tuning. Also in this case validation was used as the criterion for selecting the most suitable pretrained model; a comparison of most relevant ones is presented in Table 5 which includes model parameters and preprocessing configurations. The corresponding validation loss curves are shown in Figure 19 and Figure 18.

For each oversampling strategy, separated by color in the aforementioned table, at least one model with foreground-based normalization and one model with whole-image standardization was evaluated. Only one model with 16 × 16 patch size was reported, as any other hyperparameter configuration was consistently outperformed by the 8 × 8 ones.

Similar to the previous nomenclature, model names follow the structure: [**AugmentationType]_[StandardizationStrategy]_[Configuration]**. As for augmentation, "Base" stands for no oversampling, "Rot" stands for rotation-based augmentation and "Repl" for replication-based augmentation. As in the previous case, F and W respectively refer to foreground-based and whole-image standardization, while the configuration refers to hyperparameters.

Similarly to the anomaly detection task, foreground-based standardization proved to worsen results across configurations. Also replication-based augmentation does not provide consistent improvements, leading to the highest error values when associated to foreground standardization. On the other hand, rotation-based oversampling seems to be associated to better results.

The best performance overall is achieved by model Rot_W_2 (MSE = 0.1687), which uses rotation-based augmentations on the minority class and a reduced weight decay (min LR = 5e-09). This model was therefore selected for embedding analysis and subsequent fine-tuning.

Table 5. Most relevant pretrained models, along with hyperparameters, preprocessing steps and minimum validation loss. Each color represents a different oversampling strategy.

Model	Patch size	Data standardization	Augmentation	LR	Min LR	Min. validation loss
Base_W_0	*16 × 16*	*Whole image*	*No*	*1e-5*	*1e-08*	*0.2761*
Base_W_1	*8 × 8*	*Whole image*	*No*	*0.001*	*1e-08*	***0.1753***
Base_F_1	*8 × 8*	*Foreground*	*No*	*0.001*	*1e-08*	*0.2787*
Rot_W_1	*8 × 8*	*Whole image*	*Rotations*	*0.001*	*1e-08*	*0.1719*
Rot_W_2	*8 × 8*	*Whole image*	*Rotations*	*0.001*	*5e-09*	***0.1687***
Rot_F_1	*8 × 8*	*Foreground*	*Rotations*	*0.001*	*1e-08*	*0.2942*
Repl_W_1	*8 × 8*	*Whole image*	*Replications*	*0.001*	*1e-08*	***0.1707***
Repl_F_1	*8 × 8*	*Foreground*	*Replications*	*0.001*	*1e-08*	*0.2975*

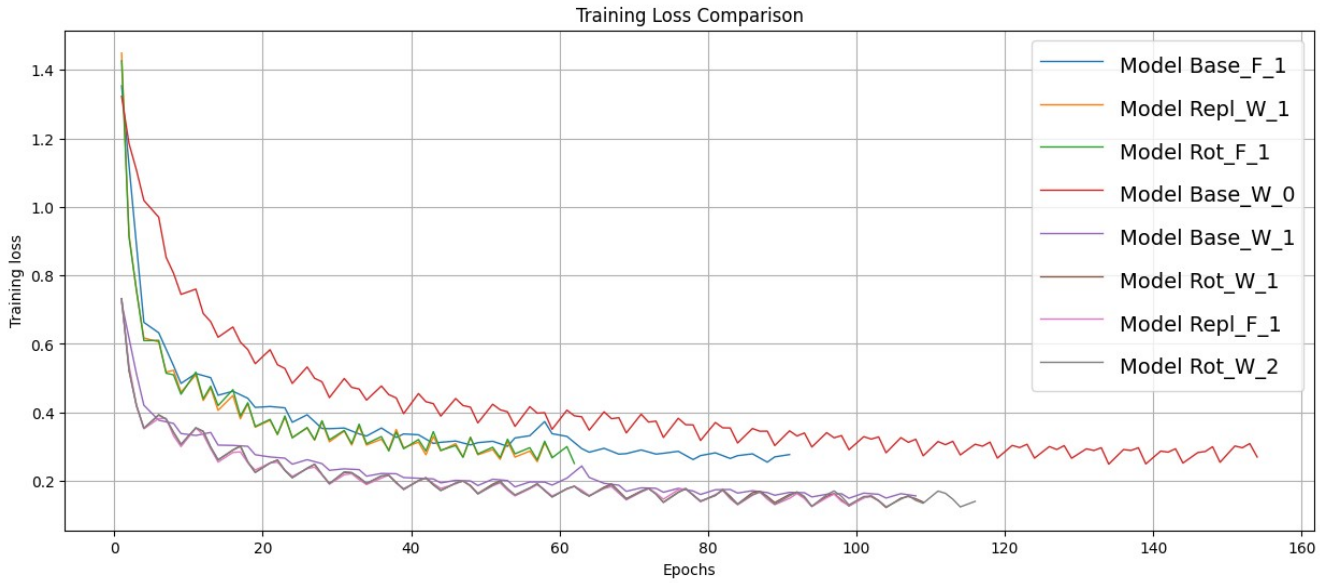


Figure 18. Comparison of training losses' curves of the most relevant models. The training loss is represented across number of epochs.

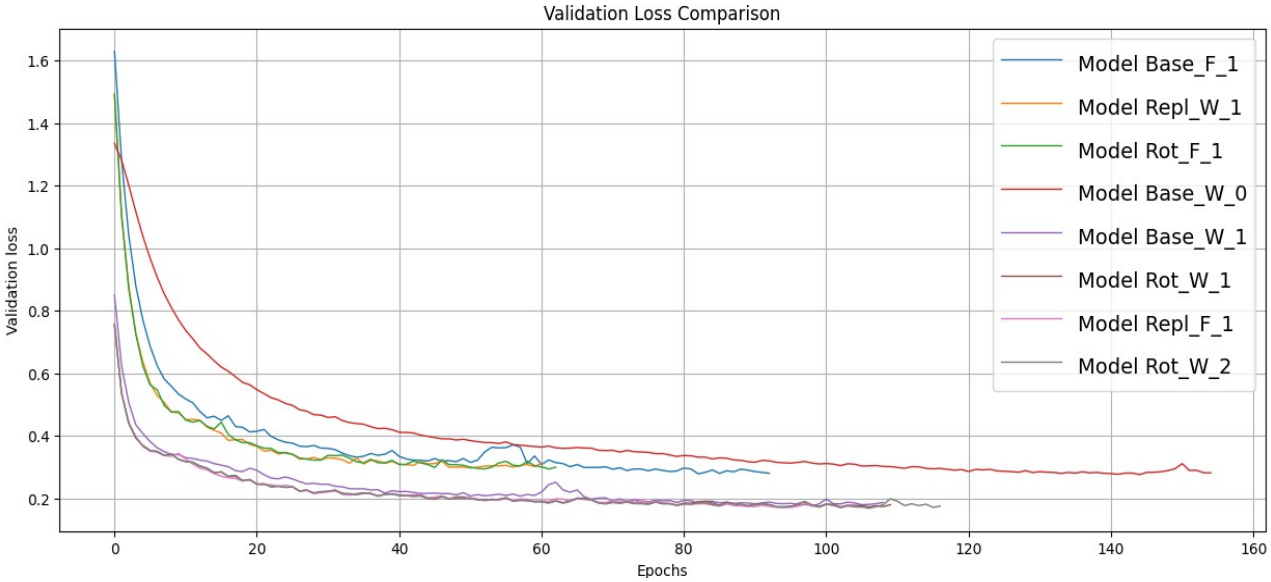


Figure 19. Comparison of validation losses' curves of the most relevant models. The validation loss is represented across the number of epochs.

3.2.2. PCA and UMAP analysis

In order to determine whether the learnt representation space already includes discriminative information that allows to distinguish between the two 30-week outcomes, PCA and UMAP techniques were applied to both CLS tokens and average embeddings of the selected model, prior to fine-tuning. Due to the limited size of the test set, the results will be presented for both training and test sets.

Training set analysis

The subplots related to the training set are shown in Figure 20.

The PCA projection of the average embeddings shows one potential positive cluster on the right area, that includes the majority of positive cases. This could suggest that the model has acquired a representation of normality that indirectly pushes abnormal cases toward this area; however, this is not entirely consistent, because a subset of positive cases continues to be outside of this cluster. In contrast, the embeddings for the negative patients seem to be widely dispersed throughout the plot, with a lack of an identifiable grouping.

When analyzing the same embeddings using UMAP, the overall structure is much more scattered for both classes. Although UMAP is designed to maintain local relationships and capture non-linear structures, in this case it does not reveal distinct clustering patterns. Instead, both positive and negative are evenly distributed; this suggests that the separability observed in PCA may be caused primarily by global variance rather than robust clustering of abnormal samples.

A similar trend is observed when looking at the CLS tokens: the PCA projection highlights once more the presence of the previous cluster, while the UMAP visualization displays a more dispersed distribution.

Overall, these findings imply that the model has learned a representation of normal data, however it cannot create tight, well-defined clusters.

Legend. *In every plot, colors represent the reference outcomes (i.e., 30-week outcomes), with red referring to positive cases and blue to negative ones, whereas shape distinguishes the two possible intermediate outcomes, with squares indicating positive 20-week evaluations and circles negative ones.*

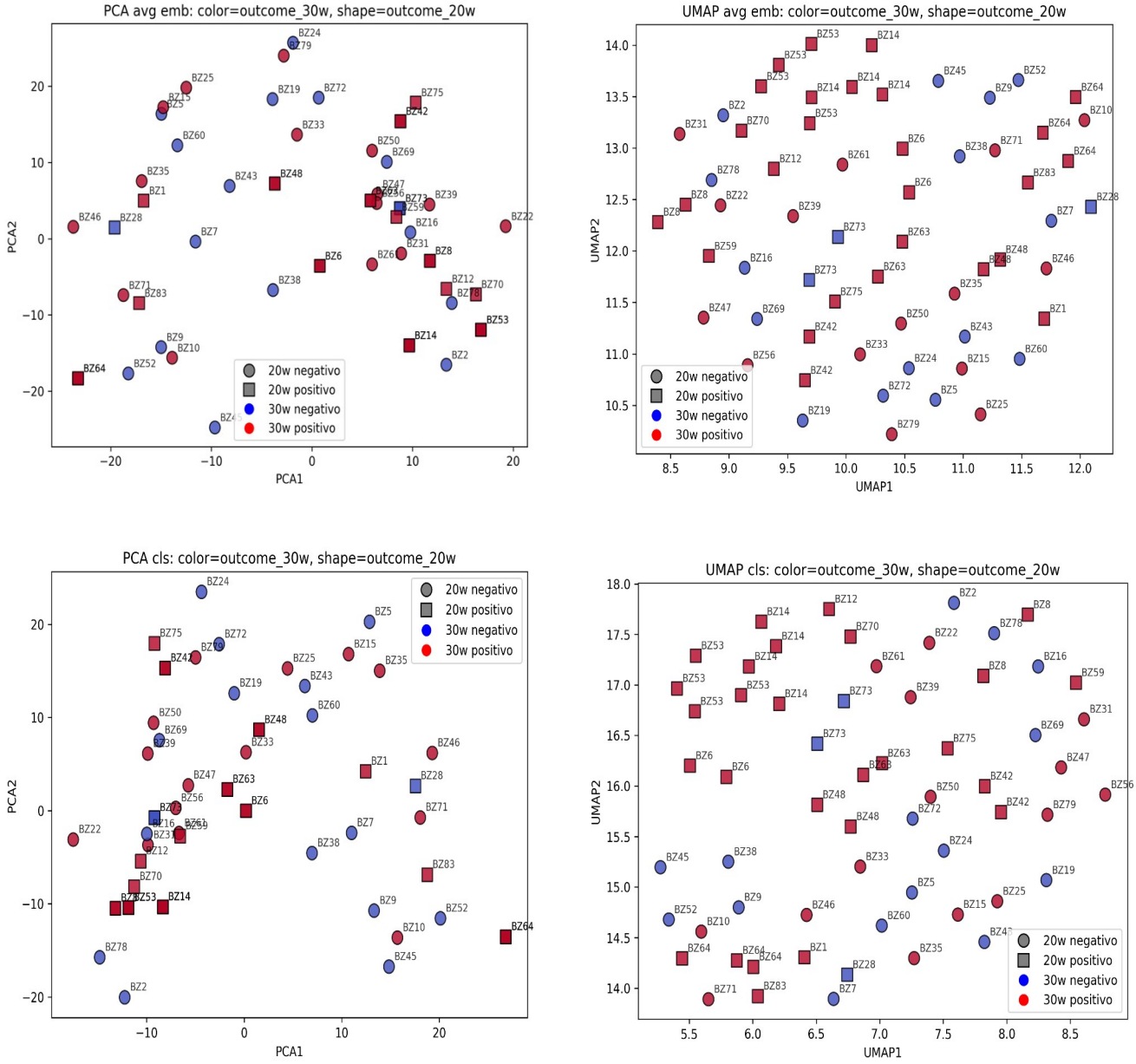


Figure 20. ***Training set.*** *Above: PCA (left) and (UMAP) of the average embedding of each patient; below: PCA (left) and UMAP (right) of the CLS token of each patient.*

Test set analysis

As for the test set, the results are illustrated in Figure 21. Overall, no significative clusters can be identified due to its limited size; the test set includes only 10 samples, consisting of 5 positives and 5 negatives, which is insufficient to reveal stable or reliable structures in the embedding space.

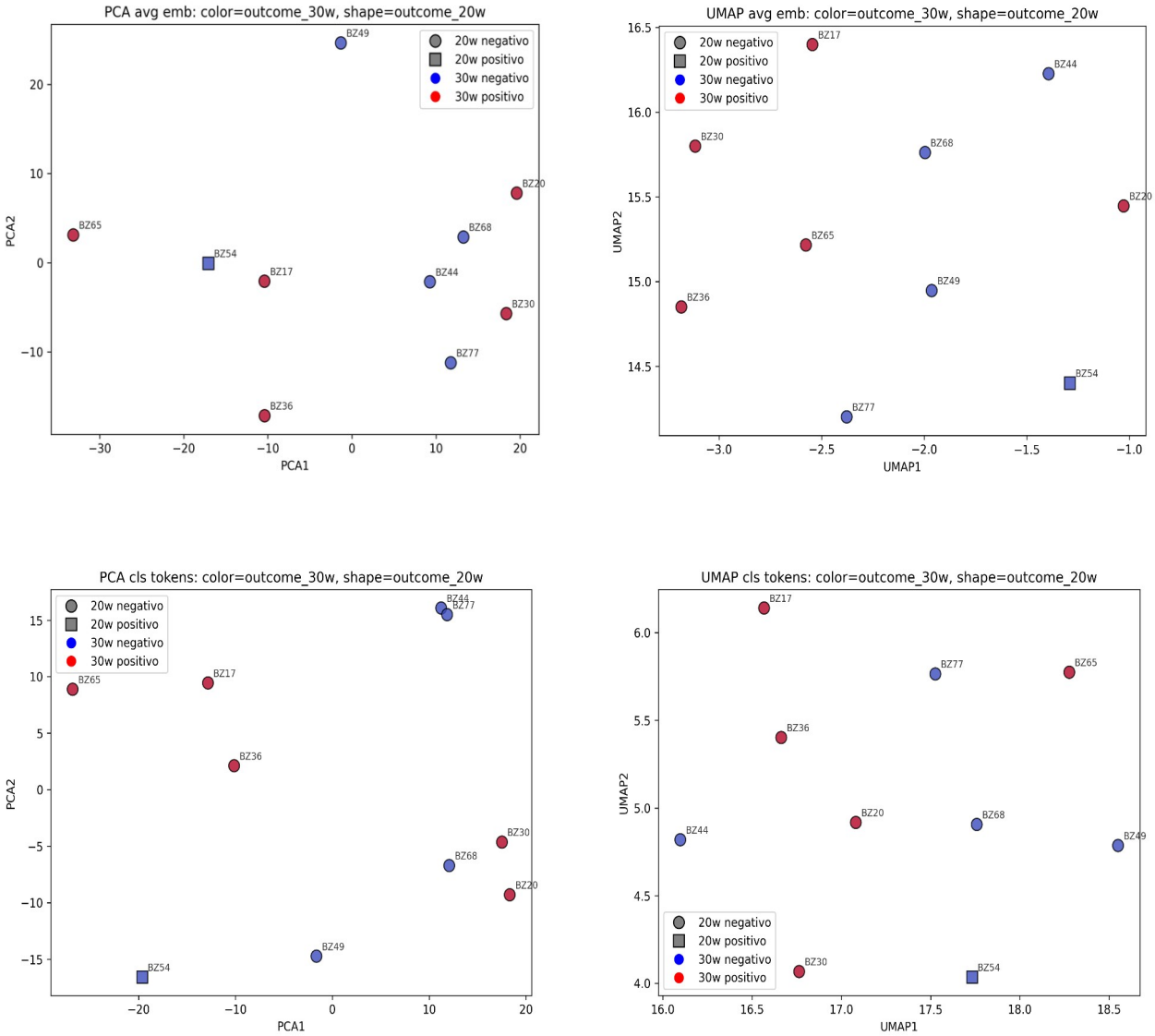


Figure 21. ***Test set.*** *Above: PCA (left) and (UMAP) of the average embedding of each patient; below: PCA (left) and UMAP (right) of the CLS token of each patient.*

3.2.3. Finetuning results

After fine-tuning the pretrained model, the final classifier was tested on the test set; however, performance was poor. Only 6 out of 10 samples were correctly identified by the classifier, which were all positives; however, further inspection of the predictions revealed that the model consistently predicted the positive class for all inputs, indicating a failure to distinguish between classes.

As a result, this model was discarded. Consequently, finetuning was applied to the other two best-performing models, one for each remaining oversampling strategy, which are Model Repl_W_1 (oversampling with replications) and Base_W_1 (no oversampling at all). Nevertheless, both models displayed the same tendency, consistently predicting only the positive class.

4. Discussion and conclusions

In order to evaluate the effectiveness of the proposed methods, this chapter provides an interpretation of the experimental findings obtained in the previous section. Additionally, the study's limitations will also be thoroughly examined, with an emphasis on the impact of data scarcity, model design and preprocessing strategies on the presented results.

Finaly, the main conclusions of the study will be outlined, along with potential directions for future research.

4.1. Findings' interpretation

4.1.1. Anomaly detection

Regarding the investigated model configurations for anomaly detection, several considerations can be drawn.

First, the use of foreground-based standardization consistently worsens performances. In this context, a plausible explanation lies in the role of background (black) pixels, which typically delineate anatomical boundaries. Particularly, forcing these values to a preset extreme (e.g., -3.0) might introduce artificial intensity distortions; this process would lead to degradation of reconstruction quality, since the model's ability to learn meaningful structural representations may be impaired.

Secondly, careful considerations have to be made on the role of patch size in ViT-MAE models. Smaller patches better retain local information, giving the model a more detailed representation of the image and therefore reduce the difficulty of the reconstruction task at the patch level. However, smaller patches also come with significant drawbacks, as they increase computational complexity and memory consumption, while also raising the danger of overfitting, since the model can become overly sensitive to local patterns and noise rather than global structures.

In the present study, larger patch configurations often exhibited poorer performances, although other factors had a significant impact on the final outcome. Nevertheless, the qualitative analysis (see Figure 10) clearly demonstrates that the 8×8 layout outperforms the 16×16 counterpart, capturing finer intensity variations in the reconstruction, particularly within the fetal brain. This implies that finer spatial granularity plays a critical role in modeling anatomically relevant details.

Thus, these observations may prompt the question of whether even smaller patches (e.g. 4×4) could further enhance performance. Although this is a reasonable hypothesis, additional experiments in this area were not feasible due to hardware constraints, particularly inadequate memory capacity; future work could therefore include the investigation of this aspect.

According to the reconstruction results, it emerges that the highest reconstruction errors are predominantly linked to the fetal region, including the placenta. This can be explained by this area's high heterogeneity in intensity distributions, which makes precise reconstruction more difficult. In contrast, the current error computation does not take into account the influence of background regions; as a result, reconstructions with larger background areas are often incorrectly favored, as the background – being relatively uniform and easier to reconstruct – falsely reducing the overall error. Therefore, the evaluation becomes biased, since background effectively dominates the reconstruction metric. Furthermore, high errors are often found also in the placenta, which could additionally diverge from diagnostical interest.

Despite this drawback, according to qualitative inspection of the reconstructed images, ViT-MAE models appear to be a promising method. In particular, the models show that they can reconstruct structural boundaries and capture differences across regions with different intensity profiles. These findings imply that the models are capable of learning meaningful representations of the underlying anatomy.

In this context, the integration of brain segmentation techniques seems essential. By isolating the region of interest (ROI) and excluding background pixels, it would be possible

to reduce the previously indicated bias and enable the model to focus on anatomically relevant structures.

These elements, combined with the limited size of the training dataset, may explain the distribution of reconstruction errors reported in Figure 17. In particular, the expected separation between positive and negative classes is not clearly observed. Even though some differences in the distributions' shape can be identified, surprisingly, higher-magnitude mistakes are frequently linked to the negative class.

As previously discussed, this behavior can result from suboptimal learning, which is probably impacted influenced by the presence of background noise and the lack of ROI-focused training. Furthermore, it is possible that the subset of negative samples in the test set was particularly unfavorable, for example due to the presence of smaller background regions compared to the positive subset, which would result to larger reconstruction errors for this class.

These findings suggest that, in the current experimental setting, reconstruction error alone may not be sufficient to robustly discriminate between the two classes, particularly without the application of fetal brain segmentation. However, these observations should be treated with caution due to the small dataset.

4.1.2. Classification task

As for the classification task, the pretraining phase validated some findings of the previous approach and provided additional relevant insights.

Coherent with the findings of the anomaly detection strategy, foreground-based normalization did not enhance model performance. This further implies that such preprocessing may be suboptimal, likely due to the distortion of boundaries' intensity as discussed previously.

In contrast, whole-image standardization produced better outcomes, especially when combined with oversampling techniques. Both identical sample replication and rotation-based augmentation improved performance, with the latter producing the best-performing model. This result is coherent with the dataset's intrinsic variability: unlike adult MRI acquisitions, fetal MRIs are characterized by considerable variations in orientation caused by fetal motion. For this reason, rotation-based augmentation provides a more realistic approximation of the underlying data distribution and improves the model's ability to generalize.

Consequently, this could lead to question why additional transformations were not introduced. In this regard, it is essential to note that some augmentations – especially those involving geometric distortions – may generate anatomically implausible patterns that potentially resemble lesions. This would introduce false artificial signals during training and ultimately reduce model reliability.

From a training dynamics' perspective, the validation loss curves show a decreasing trend, indicating that the model is capable of learning meaningful representations. This is a positive signal, as it suggests that expanding the dataset to include more participants could improve performance since it would increase anatomical diversity.

Furthermore, the integration of brain segmentation can be once again confirmed as a critical component. Even if the loss formulation differs from the anomaly detection framework – since the model is evaluated only on the reconstruction of missing patches, and does not include performance on visible ones – the reconstruction error remains strongly influenced by the presence of the background regions. As previously discussed, background areas are easier to reconstruct and therefore bias the loss values; therefore, the absence of a ROI extraction step probably contributed to the observed limited performance, since the model is forced to learn also irrelevant background information.

These limitations are also reflected in the learned representations. When the embeddings are projected into lower-dimensional spaces using PCA and UMAP, there is no clear separation between classes. A fraction of positive samples shows a weak clustering tendency

in the PCA space, which supposedly better preserves global structure, while negative samples appear more uniformly distributed. Analysis of CLS tokens shows a similar trend, which suggests that the model struggles to learn discriminative features that can successfully separate the two classes.

As a direct consequence, the fine-tuned models fail to achieve adequate classification performance. Actually, all three fine-tuned models collapse to a dysfunctional solution, which consists of predicting only the positive class. This behavior aligns with the previous findings and highlights a systemic limitation as opposed to an architecture-related problem.

Importantly, these findings suggest that the main bottleneck is not just the limited number of training samples, but also the influence of background noise on the evaluation metrics, which biases performance assessment and obstruct proper assessment of the proposed architecture.

4.2. Limitations of this study

The proposed study presents several limitations.

First, the presented evaluations were carried out on small test sets; therefore, all results should be interpreted with caution, since they may be affected by high variance and poor representativeness of data distribution, as discussed in relation to Figure 17.

Second, these findings are based exclusively on the axial plane of the SSH-TSE T2-weighted sequence. This is a clear limitation, as in clinical practice abnormalities' assessment typically relies on multiple sequences and imaging planes, which are often analyzed together to confirm the presence of hyperintensities or other relevant clinical findings across the full volumetric context.

Additionally, the selection of the best series was not performed by expert clinicians, which could have introduced potential errors or inconsistencies in the data curation process.

Furthermore, the binary labeling system for the outcomes could be overly restrictive and error-prone as well, because it may not accurately reflect the clinicians' diagnosis. In particular, otherwise physiological observations – such as below-range morphological parameters in relation to gestational age – could be incorrectly classified as positive under the proposed strategy, resulting in label noise; in such circumstances, a skilled clinician might instead assign a negative label.

Overall, these considerations highlight the need for further refinement of the dataset, labeling strategy and series' selection in order to achieve a more thorough and comprehensive evaluation of the potential of ViT-MAE models for lesion detection and classification.

4.3. Discussion of research objectives

The main objectives of this thesis were the evaluation of the potential of ViT-MAE models in predicting 30-week outcomes from 20-week MRI scans in the context of CMV infections. This was done using two AI-based strategies, an anomaly detection task and a classification task. Additionally, an important aim of this pilot study was the identification of potential limitations, in order to guide future research in this area.

While the anomaly detection method was expected to highlight abnormal regions by assigning higher reconstruction errors to samples deviating from learned normal anatomical patterns, the results only partially support this assumption. In fact, the observable differences in error magnitude distributions between the two classes are heavily impacted by the lack of segmentation, which leads to the inclusion of background noise in the reconstruction process and consequently skews the error estimation.

A similar pattern emerges from the classification task, which fails to properly discriminate the two classes as well and ultimately collapse into a degenerate classifier, that consistently predicts the majority (positive) class. This behavior was expected, since neither the learned

embeddings nor CLS token representations showed compact or well-separate clusters in the latent space.

Nevertheless, the decreasing trend in validation loss suggests that the model is capable of learning valuable representations. Even though it does not always capture fine-grained anatomical details, it is able to reconstruct global structures, edges, and overall intensity distributions with reasonable accuracy. This shows that, under appropriate conditions, the ViT-MAE models do have potential for the stated goal.

Additionally, this pilot study has revealed the beneficial impact of data augmentation strategies, raising the possibility that a larger and more diverse dataset could further enhance performance.

As outlined in the objectives, several limitations of the proposed approach were highlighted, most notably the lack of brain segmentation, which seems to be a primary cause for suboptimal performance. Finally, this analysis underscores additional sources of uncertainty that could have added noise and inconsistency to the dataset, such as the absence of expert clinical validation in both the binary labeling process and the visual inspection for scan selection.

4.4. Future work

The presented findings provide multiple directions for future work, aimed at improving both generalization performance and robustness of the model.

The encouraging results at a global level suggest that ViT-MAE models are suitable for the proposed goals, although they are strongly affected by the presence of background noise. Therefore, a primary direction for future work is the incorporation of brain segmentation techniques, which may suppress unnecessary regions and enable a more accurate recovery of fine-grained anatomical details within the fetal brain. In this regard, the FeTA challenges [52] serve as a valuable benchmark, as they offer frameworks for fetal brain segmentation tasks.

The introduction of brain segmentation would also enable the use of larger patch size, which would lower computational complexity as well as the number of tokens processed by the transformer. This would lead to increased efficiency while preserving or even enhancing performance. Additionally, multiscale architectures that integrates features extracted at different patch sizes (e.g., 16×16 and 8×8) could be explored, enabling the model to simultaneously capture information at both the global and local levels.

On the methodological side, alternative loss functions may be investigated to lessen the bias caused by intensity differences between background and foreground regions. In particular, bright-intensity regions are sometimes disproportionally penalized by typical reconstruction losses. Thus, designing loss functions that explicitly lessen the impact of background areas, or that reweight reconstruction errors based on anatomical relevance (even within the fetal brain itself), could therefore improve training stability and interpretability.

Secondly, extending the dataset to a larger and more diverse cohort would be essential to increase anatomical variability and improve model generalization. This would also potentially lead to the inclusion of 3 T MRI scans, which are increasingly common due to their better image-quality.

Finally, in order to better align the framework with clinical practice, where diagnosis is usually based on multimodal and multiplanar assessment, future work could expand the analysis to include additional imaging planes and sequences. Furthermore, the integration of complementary clinical information – such as laboratory data and ultrasound reports – could enable the development of a multimodal predictive model. Such approach could improve diagnostic accuracy and provide a more comprehensive risk estimation. In this context, replacing late gestational outcome labels with postnatal outcomes would result in more trustworthy labels, since 30-week adverse findings do not always translate into a poor postnatal prognosis. Even more importantly, the incorporation of long-term follow-up data (e.g., up to 5-6 years) would enable the detection of cases with late-onset neurodevelopmental sequelae, offering a more clinically significant evaluation framework.

Acknowledgements

I would like to sincerely thank Dr. Enea Parimbelli for his supervision in this project, and PhD candidate Lorenzo Peracchio for his major contribution to the development and implementation of the model's architecture.

I would also like to extend my sincere gratitude to the clinicians who contributed their time and expertise to this project. In particular, I would like to thank Dr. Stefano Ghirardello and Dr. Luisa Chiapparini, respectively from the Neonatal Intensive Care Unit (NICU) and the Neuroradiology Department at Fondazione IRCCS Policlinico San Matteo (Pavia), as well as Dr. Chiara Doneda from the Pediatric Radiology and Neuroradiology Unit at Vittore Buzzi Children's Hospital (Milan), since they provided extremely valuable support in interpreting medical data and defining the key clinical priorities. Their collaboration helped bridge the gap between technical development and real-world healthcare needs.

Finally, I would also like to thank Dr. Emanuela Zannin, research fellow from the Department of Neonatology and Obstetrics at Fondazione IRCCS San Gerardo dei Tintori (Monza), particularly for suggesting relevant literature and collaborating on design choices.

References

[1] K. F. M. Pontes *et al.*, 'Cytomegalovirus and Pregnancy: A Narrative Review', *J. Clin. Med.*, vol. 13, no. 2, p. 640, Jan. 2024, doi: 10.3390/jcm13020640.

[2] S. Landolfo, M. Gariglio, G. Gribaudo, and D. Lembo, 'The human cytomegalovirus', *Pharmacol. Ther.*, vol. 98, no. 3, pp. 269–297, Jun. 2003, doi: 10.1016/S0163-7258(03)00034-2.

[3] M. C. Diogo, S. Glatter, J. Binder, H. Kiss, and D. Prayer, 'The MRI spectrum of congenital cytomegalovirus infection', *Prenat. Diagn.*, vol. 40, no. 1, pp. 110–124, Jan. 2020, doi: 10.1002/pd.5591.

[4] A. Y. Yamamoto *et al.*, 'Congenital Cytomegalovirus Infection as a Cause of Sensorineural Hearing Loss in a Highly Immune Population', *Pediatr. Infect. Dis. J.*, vol. 30, no. 12, p. 1043, Dec. 2011, doi: 10.1097/INF.0b013e31822d9640.

[5] Z. Shang and X. Li, 'Human cytomegalovirus: pathogenesis, prevention, and treatment', *Mol. Biomed.*, vol. 5, p. 61, Nov. 2024, doi: 10.1186/s43556-024-00226-7.

[6] M. H. Pesch and M. R. Schleiss, 'Emerging Concepts in Congenital Cytomegalovirus', *Pediatrics*, vol. 150, no. 2, Aug. 2022, doi: 10.1542/peds.2021-055896.

[7] E. C. Swanson and M. R. Schleiss, 'Congenital Cytomegalovirus Infection: New Prospects for Prevention and Therapy', *Pediatr. Clin. North Am.*, vol. 60, no. 2, pp. 335–349, Apr. 2013, doi: 10.1016/j.pcl.2012.12.008.

[8] I. N. Pardieck, G. Beyrend, A. Redeker, and R. Arens, 'Cytomegalovirus infection and progressive differentiation of effector-memory T cells', *F1000Research*, vol. 7, p. F1000 Faculty Rev-1554, 2018, doi: 10.12688/f1000research.15753.1.

[9] T. M. Lanzieri, S. C. Dollard, S. R. Bialek, and S. D. Grosse, 'Systematic review of the birth prevalence of congenital cytomegalovirus infection in developing countries', *Int. J. Infect. Dis. IJID Off. Publ. Int. Soc. Infect. Dis.*, vol. 22, pp. 44–48, May 2014, doi: 10.1016/j.ijid.2013.12.010.

[10] G. Singh and A. Gaidhane, 'A Review of Sensorineural Hearing Loss in Congenital Cytomegalovirus Infection', *Cureus*, Oct. 2022, doi: 10.7759/cureus.30703.

[11] S. Salomè *et al.*, 'Congenital cytomegalovirus infection: the state of the art and future perspectives', *Front. Pediatr.*, vol. 11, Nov. 2023, doi: 10.3389/fped.2023.1276912.

[12] C. Chatzakis, Y. Ville, G. Makrydimas, K. Dinas, A. Zavlanos, and A. Sotiriadis, 'Timing of primary maternal cytomegalovirus infection and rates of vertical transmission and fetal consequences', *Am. J. Obstet. Gynecol.*, vol. 223, no. 6, pp. 870-883.e11, Dec. 2020, doi: 10.1016/j.ajog.2020.05.038.

[13] F. Scaramuzzino *et al.*, 'Secondary cytomegalovirus infections: How much do we still not know? Comparison of children with symptomatic congenital cytomegalovirus born to mothers with primary and secondary infection', *Front. Pediatr.*, vol. 10, Jul. 2022, doi: 10.3389/fped.2022.885926.

[14] M. Leruez-Ville *et al.*, 'Risk Factors for Congenital Cytomegalovirus Infection Following Primary and Nonprimary Maternal Infection: A Prospective Neonatal

Screening Study Using Polymerase Chain Reaction in Saliva', *Clin. Infect. Dis.*, vol. 65, no. 3, pp. 398–404, Aug. 2017, doi: 10.1093/cid/cix337.
[15] M. Leruez-Ville *et al.*, 'Quantifying the Burden of Congenital Cytomegalovirus Infection With Long-term Sequelae in Subsequent Pregnancies of Women Seronegative at Their First Pregnancy', *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.*, vol. 71, no. 7, pp. 1598–1603, Oct. 2020, doi: 10.1093/cid/ciz1067.
[16] K. B. Fowler, S. Stagno, R. F. Pass, W. J. Britt, T. J. Boll, and C. A. Alford, 'The outcome of congenital cytomegalovirus infection in relation to maternal antibody status', *N. Engl. J. Med.*, vol. 326, no. 10, pp. 663–667, Mar. 1992, doi: 10.1056/NEJM199203053261003.
[17] M. Rybak-Krzyszkowska *et al.*, 'Cytomegalovirus Infection in Pregnancy Prevention and Treatment Options: A Systematic Review and Meta-Analysis', *Viruses*, vol. 15, no. 11, p. 2142, Nov. 2023, doi: 10.3390/v15112142.
[18] M. C.-J. Cheeran, J. R. Lokensgard, and M. R. Schleiss, 'Neuropathogenesis of Congenital Cytomegalovirus Infection: Disease Mechanisms and Prospects for Intervention', *Clin. Microbiol. Rev.*, vol. 22, no. 1, pp. 99–126, Jan. 2009, doi: 10.1128/cmr.00023-08.
[19] G. Piccirilli *et al.*, 'Fetal Brain Damage in Human Fetuses with Congenital Cytomegalovirus Infection: Histological Features and Viral Tropism', *Cell. Mol. Neurobiol.*, vol. 43, no. 3, pp. 1385–1399, Apr. 2023, doi: 10.1007/s10571-022-01258-9.
[20] S. E. Luck *et al.*, 'Congenital Cytomegalovirus: A European Expert Consensus Statement on Diagnosis and Management', *Pediatr. Infect. Dis. J.*, vol. 36, no. 12, p. 1205, Dec. 2017, doi: 10.1097/INF.0000000000001763.
[21] A. Chiereghin *et al.*, 'Universal Newborn Screening for Congenital Cytomegalovirus Infection – From Infant to Maternal Infection: A Prospective Multicenter Study', *Front. Pediatr.*, vol. 10, Jul. 2022, doi: 10.3389/fped.2022.909646.
[22] P.-G. Maltezou *et al.*, 'Maternal type of CMV infection and sequelae in infants with congenital CMV: Systematic review and meta-analysis', *J. Clin. Virol.*, vol. 129, p. 104518, Aug. 2020, doi: 10.1016/j.jcv.2020.104518.
[23] I. Foulon, A. Naessens, W. Foulon, A. Casteels, and F. Gordts, 'A 10-year prospective study of sensorineural hearing loss in children with congenital cytomegalovirus infection', *J. Pediatr.*, vol. 153, no. 1, pp. 84–88, Jul. 2008, doi: 10.1016/j.jpeds.2007.12.049.
[24] A. Giannattasio *et al.*, 'Outcomes of congenital cytomegalovirus disease following maternal primary and non-primary infection', *J. Clin. Virol. Off. Publ. Pan Am. Soc. Clin. Virol.*, vol. 96, pp. 32–36, Nov. 2017, doi: 10.1016/j.jcv.2017.09.006.
[25] M. Leruez-Ville *et al.*, 'Consensus recommendation for prenatal, neonatal and postnatal management of congenital cytomegalovirus infection from the European congenital infection initiative (ECCI)', *Lancet Reg. Health – Eur.*, vol. 40, May 2024, doi: 10.1016/j.lanepe.2024.100892.
[26] T. Lazzarotto *et al.*, 'Congenital Cytomegalovirus Infection: A Narrative Review of the Issues in Screening and Management From a Panel of European Experts', *Front. Pediatr.*, vol. 8, Jan. 2020, doi: 10.3389/fped.2020.00013.

[27] C. Périllaud-Dubois *et al.*, 'Cost-effectiveness of screening and valacyclovir-based treatment strategies for first-trimester cytomegalovirus primary infection in pregnant women in France', *Ultrasound Obstet. Gynecol.*, vol. 62, no. 4, pp. 573–584, 2023, doi: 10.1002/uog.26226.
[28] G. H. Shim, 'Treatment of congenital cytomegalovirus infection', *Clin. Exp. Pediatr.*, vol. 66, no. 9, pp. 384–394, Dec. 2022, doi: 10.3345/cep.2022.01032.
[29] M. Leruez-Ville *et al.*, 'Consensus recommendation for prenatal, neonatal and postnatal management of congenital cytomegalovirus infection from the European congenital infection initiative (ECCI)', *Lancet Reg. Health – Eur.*, vol. 40, May 2024, doi: 10.1016/j.lanepe.2024.100892.
[30] S. N. Saleem, 'Fetal MRI: An approach to practice: A review', *J. Adv. Res.*, vol. 5, no. 5, pp. 507–523, Sep. 2014, doi: 10.1016/j.jare.2013.06.001.
[31] R.-J. M. van Geuns *et al.*, 'Basic principles of magnetic resonance imaging', *Prog. Cardiovasc. Dis.*, vol. 42, no. 2, pp. 149–156, Sep. 1999, doi: 10.1016/S0033-0620(99)70014-9.
[32] N. Mohindra and Z. Neyaz, 'Magnetic resonance sequences: Practical neurological applications', *Neurol. India*, vol. 63, no. 2, p. 241, Apr. 2015, doi: 10.4103/0028-3886.156293.
[33] L. Manganaro *et al.*, 'Fetal MRI: what's new? A short review', *Eur. Radiol. Exp.*, vol. 7, no. 1, p. 41, Aug. 2023, doi: 10.1186/s41747-023-00358-5.
[34] C. Doneda *et al.*, 'Early Cerebral Lesions in Cytomegalovirus Infection: Prenatal MR Imaging', *Radiology*, vol. 255, no. 2, pp. 613–621, May 2010, doi: 10.1148/radiol.10090749.
[35] R. Birnbaum, L. Ben-Sira, T. Lerman-Sagie, and G. Malinger, 'The use of fetal neurosonography and brain MRI in cases of cytomegalovirus infection during pregnancy: A retrospective analysis with outcome correlation', *Prenat. Diagn.*, vol. 37, no. 13, pp. 1335–1342, Dec. 2017, doi: 10.1002/pd.5180.
[36] G. Lucignani *et al.*, 'A new MRI severity score to predict long-term adverse neurologic outcomes in children with congenital Cytomegalovirus infection', *J. Matern.-Fetal Neonatal Med. Off. J. Eur. Assoc. Perinat. Med. Fed. Asia Ocean. Perinat. Soc. Int. Soc. Perinat. Obstet.*, vol. 34, no. 6, pp. 859–866, Mar. 2021, doi: 10.1080/14767058.2019.1620725.
[37] R. Calandrelli *et al.*, 'Visual-semiquantitative and quantitative magnetic resonance imaging (MRI) data may predict outcome in congenital cytomegalovirus infection', *Clin. Radiol.*, vol. 88, Sep. 2025, doi: 10.1016/j.crad.2025.106991.
[38] A. Haleem, M. Javaid, and I. H. Khan, 'Current status and applications of Artificial Intelligence (AI) in medical field: An overview', *Curr. Med. Res. Pract.*, vol. 9, no. 6, pp. 231–237, Nov. 2019, doi: 10.1016/j.cmrp.2019.11.005.
[39] Z. Feng and X. Hua, 'Applications and Current Status of AI in the Medical Field', *J. Phys. Conf. Ser.*, vol. 2289, no. 1, p. 012030, Jun. 2022, doi: 10.1088/1742-6596/2289/1/012030.
[40] K. Kaur, C. Singh, and Y. Kumar, 'Diagnosis and Detection of Congenital Diseases in New-Borns or Fetuses Using Artificial Intelligence Techniques: A Systematic Review',

Arch. Comput. Methods Eng., vol. 30, no. 5, pp. 3031–3058, Jun. 2023, doi: 10.1007/s11831-023-09892-2.

[41] B. Hunter, S. Hindocha, and R. W. Lee, 'The Role of Artificial Intelligence in Early Cancer Diagnosis', *Cancers*, vol. 14, no. 6, p. 1524, Mar. 2022, doi: 10.3390/cancers14061524.

[42] A. P. Hederman *et al.*, 'Discrimination of primary and chronic cytomegalovirus infection based on humoral immune profiles in pregnancy', *J. Clin. Invest.*, vol. 134, no. 20, Oct. 2024, doi: 10.1172/JCI180560.

[43] R. Arav-Boger, Y. S. Boger, C. B. Foster, and Z. Boger, 'The Use of Artificial Neural Networks in Prediction of Congenital CMV Outcome from Sequence Data', *Bioinforma. Biol. Insights*, vol. 2, p. BBI.S764, Jan. 2008, doi: 10.4137/BBI.S764.

[44] K. Annelies *et al.*, 'Results of a multicenter registry for congenital cytomegalovirus infection in Flanders, Belgium: From prenatal diagnosis over neonatal management to therapy', *Early Hum. Dev.*, vol. 163, p. 105499, Dec. 2021, doi: 10.1016/j.earlhumdev.2021.105499.

[45] K. He, X. Chen, S. Xie, Y. Li, P. Dollar, and R. Girshick, 'Masked Autoencoders Are Scalable Vision Learners', in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA: IEEE, Jun. 2022, pp. 15979–15988. doi: 10.1109/CVPR52688.2022.01553.

[46] A. C. Erdur *et al.*, 'MultiMAE for Brain MRIs: Robustness to Missing Inputs Using Multi-modal Masked Autoencoder', in *Machine Learning in Medical Imaging: 16th International Workshop, MLMI 2025, Held in Conjunction with MICCAI 2025, Daejeon, South Korea, September 23, 2025, Proceedings*, Berlin, Heidelberg: Springer-Verlag, Sep. 2025, pp. 572–582. doi: 10.1007/978-3-032-09513-8_55.

[47] M. H. Sikder, S. T. Sharif, M. S. Hossain Shovon, J. Shin, and M. F. Mridha, 'Weighted Loss Integrated Fine-Tuned ViT Model for Multi-Class Human Facial Emotion Recognition', in *2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, Nov. 2024, pp. 487–494. doi: 10.1109/3ict64318.2024.10824313.

[48] A. Gupta, I. Osman, M. S. Shehata, W. J. Braun, and R. E. Feldman, 'MedMAE: A Self-Supervised Backbone for Medical Imaging Tasks', *Computation*, vol. 13, no. 4, p. 88, Apr. 2025, doi: 10.3390/computation13040088.

[49] 'Multispectral 3D Masked Autoencoders for Anomaly Detection in Non-Contrast Enhanced Breast MRI - Google Search'. Accessed: Apr. 07, 2026. [Online]. Available: https://link.springer.com/chapter/10.1007/978-3-031-45350-2_5

[50] O. A. Glenn, 'MR imaging of the fetal brain', *Pediatr. Radiol.*, vol. 40, no. 1, pp. 68–81, 2010, doi: 10.1007/s00247-009-1459-3.

[51] Z. Zhao *et al.*, 'Automatic detection of the fetal brain midsagittal plane on MRI using a deep learning pipeline', *Am. J. Obstet. Gynecol. MFM*, p. 101962, Apr. 2026, doi: 10.1016/j.ajogmf.2026.101962.

[52] V. Zalevskyi *et al.*, 'Advances in automated fetal brain MRI segmentation and biometry: Insights from the FeTA 2024 challenge', *Med. Image Anal.*, vol. 109, p. 103941, Mar. 2026, doi: 10.1016/j.media.2026.103941.